# A large deviation approach to optimal transport

Christian Léonard

# A LARGE DEVIATION APPROACH TO OPTIMAL TRANSPORT

CHRISTIAN LÉONARD

Abstract. A probabilistic method for solving the Monge-Kantorovich mass transport problem on $\mathbb{R}^d$ is introduced. A system of empirical measures of independent particles is built in such a way that it obeys a doubly indexed large deviation principle with an optimal transport cost as its rate function. As a consequence, new approximation results for the optimal cost function and the optimal transport plans are derived. They follow from the $\Gamma$-convergence of a sequence of normalized relative entropies toward the optimal transport cost. A wide class of cost functions including the standard power cost functions $|x-y|^p$ enter this framework.

## 1. Introduction

This paper introduces a probabilistic method for solving the Monge-Kantorovich mass transport problem.

1.1. **The Monge-Kantorovich problem.** Let $\mu$ and $\nu$ be two probability measures on $\mathbb{R}^d$ seen as mass distributions. One wants to transfer $\mu$ to $\nu$ with a minimal cost, given that transporting a unit mass from $x_0$ to $x_1$ costs $c(x_0, x_1)$. This means that one searches for a transport plan $x_1 = T(x_0)$ such that the image measure $T \diamond \mu$ is $\nu$ and $\int_{\mathbb{R}^d} c(x_0, T(x_0))\, \mu(dx_0)$ is minimal. This problem was addressed by G. Monge [17] at the eighteenth century. In the 40's, L.V. Kantorovich [12], [13] proposed a relaxed version of Monge problem by allowing each cell of mass at $x_0$ to crumble into powder so that it can be tranfered to several $x_1$'s. In mathematical terms, one searches for a probability measure $\rho$ on $\mathbb{R}^d \times \mathbb{R}^d$ whose marginal measures $\rho_0(dx_0) = \rho(dx_0 \times \mathbb{R}^d)$ and $\rho_1(dx_1) = \rho(\mathbb{R}^d \times dx_1)$ satisfy $\rho_0 = \mu$ and $\rho_1 = \nu$ and such that $\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x_0, x_1)\, \rho(dx_0 dx_1)$ is minimal. Let us denote $\mathcal{P}_{\mathbb{R}^d}$ and $\mathcal{P}_{\mathbb{R}^{2d}}$ the sets of all probability measures on $\mathbb{R}^d$ and $\mathbb{R}^d \times \mathbb{R}^d$. For each $\mu$ and $\nu$ in $\mathcal{P}_{\mathbb{R}^d}$, we face the optimization problem

$$\text{(MK)} \qquad \text{minimize} \quad \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x_0, x_1)\, \rho(dx_0 dx_1) \quad \text{subject to} \quad \rho \in \Pi(\mu, \nu)$$

where the cost function $c : \mathbb{R}^d \times \mathbb{R}^d \mapsto [0, +\infty]$ is assumed to be measurable and

$$\Pi(\mu, \nu) = \{\rho \in \mathcal{P}_{\mathbb{R}^{2d}}; \rho_0 = \mu, \rho_1 = \nu\}$$

is the set of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $\mu$ and $\nu$. This problem is called the Monge-Kantorovich mass transport problem. Monge problem corresponds to the transport plans $\rho(dx_0 dx_1) = \mu(dx_0)\delta_{T(x_0)}(dx_1)$ where $\delta$ stands for the Dirac measure. Kantorovich's relaxation procedure embeds Monge's nonlinear problem in the linear programing problem (MK).

*Date*: December 2005.

2000 *Mathematics Subject Classification.* 49J45, 49J53, 58E99, 60F10, 60G57, 90B06.

*Key words and phrases.* Monge-Kantorovich mass transport problem, Large deviations of empirical measures, $\Gamma$-convergence, Doubly indexed large deviation principle, relative entropy.

The value of (MK) is the *transportation cost* defined for all $\mu$ and $\nu$ in $\mathcal{P}_{\mathbb{R}^d}$ by

$$\mathcal{T}_c(\mu,\nu) := \inf_{\rho\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} c(x_0,x_1)\,\rho(dx_0dx_1). \tag{1.1}$$

The special cost function $c_p(x_0,x_1) = |x_1-x_0|^p$ with $p\geq 1$, leads to the Wassertein metric $\mathcal{T}_{c_p}^{1/p}(\mu,\nu)$.

1.2. **Which large deviations?** As the title of the paper indicates, our probabilistic approach of Monge-Kantorovich problem is in terms of large deviations. One can interpret $\mu$ and $\nu$ respectively as the distributions of the initial and final random positions $X_0$ and $X_1$ of a random process $(X_t)_{0\leq t\leq 1}$. In the present paper, only the couple of initial and final positions $(X_0,X_1)$ is considered.

Our aim is to obtain a Large Deviation Principle (LDP) in $\mathcal{P}_{\mathbb{R}^d}$ the rate function of which is $\nu\mapsto \mathcal{T}_c(\mu,\nu)$ where $\mu$ is fixed. The definition of a LDP is recalled at (1.6). General cost functions will be considered in the article but for the sake of clarity, in this introductory section our procedure is described in the special case of the quadratic cost funtion $c(x_0,x_1)=|x_1-x_0|^2/2$. For each integer $k\geq 1$, take a system of $n$ independent random couples $(X^k_{n,i}(0),X^k_{n,i}(1))_{1\leq i\leq n}$ which is described as follows. For each $i$, the initial position $X^k_{n,i}(0)=z_{n,i}$ is deterministic and the final position is

$$X^k_{n,i}(1) = z_{n,i} + Y_i/\sqrt{k}$$

where the $Y_i$'s are independent copies of a standard normal vector in $\mathbb{R}^d$. Consider the initial mass distribution $\mu$ as fixed and deterministic and choose the initial positions $z_{n,i}$ in such a way that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \delta_{z_{n,i}} = \mu.$$

The empirical measure of the final positions is

$$N^k_n = \frac{1}{n}\sum_{i=1}^n \delta_{X^k_{n,i}(1)}.$$

It is a random element of $\mathcal{P}_{\mathbb{R}^d}$. An easy variation of Sanov's theorem states that for each $k$ the system $\{N^k_n\}_{n\geq 1}$ obeys the LDP in $\mathcal{P}_{\mathbb{R}^d}$ with speed $n$ and the rate function

$$\nu\in\mathcal{P}_{\mathbb{R}^d}\mapsto \inf_{\rho\in\Pi(\mu,\nu)} H(\rho|\pi^k)\in[0,\infty]. \tag{1.2}$$

Here, $H(\rho|\pi^k)$ is the relative entropy (see (2.15) for its definition) of $\rho$ with respect to $\pi^k$ and $\pi^k\in\mathcal{P}_{\mathbb{R}^{2d}}$ is the law of $(Z, Z+Y/\sqrt{k})$ where $Z$ and $Y$ are independent, the law of $Z$ is $\mu$ and $Y$ is a standard normal vector. On the other hand, $\{Y/\sqrt{k}\}_{k\geq 1}$ obeys the LDP in $\mathbb{R}^d$ as $k$ tends to infinity with speed $k$ and rate function $c(u)=|u|^2/2$.
Since

(i) the speed of the LDP for $\{Y/\sqrt{k}\}_{k\geq 1}$ is $k$ and
(ii) the rate functions (1.2) and $c(u)=|u|^2/2$ are reminiscent of $\mathcal{T}_c$ given at (1.1),

it wouldn't be surprising that

(i) the order of magnitude of $H(\rho|\pi^k)$ is $k$ and
(ii) one should mix together two types of LDPs with $n$ and $k$ tending to infinity, in order to obtain some LDP with the rate function $\nu\mapsto\mathcal{T}_c(\mu,\nu)$.

Indeed, denoting for each $\nu \in \mathcal{P}_{\mathbb{R}^d}$ with fixed $\mu$,

$$\begin{aligned} T_k(\nu) &= \inf_{\rho\in\Pi(\mu,\nu)} H(\rho|\pi^k)/k \quad \text{and} \\ T(\nu) &= \mathcal{T}_c(\mu,\nu), \end{aligned}$$

it will be proved that the following $\Gamma$-convergence result

$$\Gamma\text{-}\lim_{k\to\infty} T_k = T \tag{1.3}$$

holds. As a consequence of this convergence result, for each $\nu \in \mathcal{P}_{\mathbb{R}^d}$, there exists a sequence $(\nu_k)_{k\geq 1}$ such that

$$\lim_{k\to\infty} \nu_k = \nu \quad \text{and} \quad \lim_{k\to\infty} \inf_{\rho\in\Pi(\mu,\nu_k)} H(\rho|\pi^k)/k = \mathcal{T}_c(\mu,\nu). \tag{1.4}$$

Theorem 2.9 is the main result of the paper. It states that $\{N_n^k\}_{k,n\geq 1}$ obeys the doubly indexed LDP as $n$ first tends to infinity, then $k$ tends to infinity with speed $kn$ and rate function $T$, see Definition 2.5 for the notion of doubly indexed LDP.

1.3. **An approximation procedure.** The $\Gamma$-limit (1.3) suggests that the sequence of minimizers $\rho_k^*$ of $H(\rho|\pi^k)$ subject to the constraint $\rho \in \Pi(\mu,\nu)$ should converge as $k$ tends to infinity to some minimizer of $\rho \mapsto \int_{\mathbb{R}^{2d}} c\,d\rho$ subject to the same constraint $\rho \in \Pi(\mu,\nu)$. This fails in many situations. Consider for instance a purely atomic initial measure $\mu$ and a family of atomic probability measures $\pi^k$. Although $T(\nu)$ may be finite for some diffuse final measure $\nu$, there are no $\rho$ in $\Pi(\mu,\nu)$ which are absolutely continuous with respect to $\pi^k$ since $\pi_1^k$ is atomic. Hence, $T_k(\nu) = +\infty$ for all $k$, and there are no minimizers $\rho_k^*$ at all. To take this phenomenon into account, one can think of the minimization problems

$$(\mathrm{MK}_k) \qquad \text{minimize} \quad H(\rho|\pi^k)/k \quad \text{subject to} \quad \rho \in \Pi(\mu,\nu_k)$$

where $(\nu_k)_{k\geq 1}$ satisfies (1.4). I didn't succeed in proving that $\lim_{k\to\infty}(\mathrm{MK}_k) = (\mathrm{MK})$ in the sense of $\Gamma$-convergence.

Alternately, one can relax the constraint $\rho_1 = \nu$ by means of a continuous penalization sequence and consider the three minimization problems

$$(\mathrm{MK}_k^\alpha) \qquad \text{minimize} \quad H(\rho|\pi^k)/k + \alpha d(\rho_1,\nu) \quad \text{subject to} \quad \rho_0 = \mu$$

$$(\mathrm{MK}^\alpha) \qquad \text{minimize} \quad \int_{\mathbb{R}^{2d}} c\,d\rho + \alpha d(\rho_1,\nu) \quad \text{subject to} \quad \rho_0 = \mu$$

$$(\mathrm{MK}) \qquad \text{minimize} \quad \int_{\mathbb{R}^{2d}} c\,d\rho \quad \text{subject to} \quad \rho \in \Pi(\mu,\nu)$$

where $k, \alpha \geq 1$ are intended to tend to infinity and $d(\rho_1,\nu)$ is some distance between $\rho_1$ and $\nu$ which is compatible with the narrow topology of $\mathcal{P}_{\mathbb{R}^d}$.

Note that $(\mathrm{MK}_k^\alpha)$ is a strictly convex problem while $(\mathrm{MK}^\alpha)$ and (MK) are not. As a consequence $(\mathrm{MK}_k^\alpha)$ admits a unique minimizer $\rho_k^\alpha$ while $(\mathrm{MK}^\alpha)$ and (MK) may admit several ones. It will proved by means of another $\Gamma$-convergence result that

$$\lim_{k\to\infty} (\mathrm{MK}_k^\alpha) = (\mathrm{MK}^\alpha) \quad \text{and} \quad \lim_{\alpha\to\infty} (\mathrm{MK}^\alpha) = (\mathrm{MK}). \tag{1.5}$$

These formulas are to be understood at a formal level. It means in particular that for each $\alpha$, $\lim_{k\to\infty} \inf(\mathrm{MK}_k^\alpha) = \inf(\mathrm{MK}^\alpha)$ and all the limit points of the relatively compact sequence $(\rho_k^\alpha)_{k\geq 1}$ are minimizers of the limiting problem $(\mathrm{MK}^\alpha)$. Similarly, $\lim_{\alpha\to\infty} \inf(\mathrm{MK}^\alpha) = \inf(\mathrm{MK})$ and denoting $\rho^\alpha$ a minimizer of $(\mathrm{MK}^\alpha)$, any limit point of the relatively compact sequence $(\rho^\alpha)_{\alpha\geq 1}$ is a minimizer of the limiting problem (MK).

1.4. **Some comment about the results of this paper.** The doubly indexed LDP for $\{N_n^k\}_{k,n\geq 1}$, the limit (1.3) and the approximation procedure (1.5) are new results. Large deviations have only been used as a guideline to obtain the analytical results (1.3) and (1.5).

In the rest of the paper not only the quadratic cost is considered but a much wider class of cost functions. In particular, the above mentioned results hold true for the usual power cost functions $c(u) = |u|^p$ with $p > 0$. Note that the convexity of $c$ is not required.

We choosed $\mathbb{R}^d$ as the surrounding space to make the presentation of the results easier. It is by no way a limitation. Our main large deviation result (Theorem 5.1) is stated with Polish spaces. On the other hand, the proofs of our convergence results mainly rely on $\Gamma$-convergence. We have done them in $\mathbb{R}^d$, but their extension to a Polish space is obvious.

As a by-product of our approach, the Kantorovich duality ([24], Theorem 1.3) is recovered, see Theorems 5.1 and 6.2. This provides a new proof of it, although not the shortest one.

1.5. **Literature.** Since Brenier's note [5] in 1987 which was motivated by fluid mechanics, optimal transport is a very active area of applied mathematics. For a comprehensive account on optimal transport theory, we refer to the monographs of Rachev and Rüschendorf [19] and Villani [24]. Villani's recent Saint-Flour lecture notes [25] are up-to-date and aimed at a probalistic reader. They introduce newly born techniques and offer a very long reference list.

Although optimal transport has important consequences in probability theory (Wasserstein's metrics or transportation inequalities for instance), it has seldom been studied from a probabilistic point of view. Let us cite among others the contributions of Feyel and Üstünel [10], [11] about the Monge-Kantorovich problem on Wiener space. Recently, Mikami [16] has obtained a probabilistic proof of the existence of a solution to Monge's problem with a quadratic cost by means of an approximation procedure by $h$-processes. His approach is based on optimal control techniques.

Doubly indexed LDs of empirical measures have been studied by Boucher, Ellis and Turkington in [3]. In [14], the tight connection between doubly indexed LDs and the $\Gamma$-convergence of LD rate functions is stressed. This will be used in the present article.

1.6. **$\Gamma$-convergence.** The $\Gamma$-convergence is a useful tool which is going to be used repeatedly. We refer to the monograph of G. Dal Maso [15] for a clear exposition of the subject. Precise references to the invoked theorems in [15] will be written all along the paper.

Recall that if it exists, the $\Gamma$-limit of the sequence $(f_n)_{n\geq 1}$ of $(-\infty, \infty]$-valued functions on a topological space $X$ is given for all $x$ in $X$ by

$$\Gamma\text{-}\lim_{n\to\infty} f_n(x) = \sup_{V\in\mathcal{N}(x)} \lim_{n\to\infty} \inf_{y\in V} f_n(y)$$

where $\mathcal{N}(x)$ is the set of all neighbourhoods of $x$. This notion of convergence is well-designed for minimization problems. Denoting $f = \Gamma\text{-}\lim_{n\to\infty} f_n$ and taking $(x_n)$ a converging sequence of minimizers of $(f_n)$ with $\lim_{n\to\infty} x_n = x^*$, if $(f_n)_{n\geq 1}$ is equi-coercive we have $\lim_{n\to\infty} \inf f_n = \inf f$ and $x^*$ is a minimizer of $f$.

1.7. **Some notations and conventions.** Let us fix some notations and conventions.

*Topological conventions.* The space of all continuous bounded functions on a topological space $\mathcal{X}$ is denoted by $C_{\mathcal{X}}$ and is equipped with the uniform norm $\|f\| = \sup_{x\in\mathcal{X}} |f(x)|$, $f \in C_{\mathcal{X}}$. Unless specified, its dual space $C'_{\mathcal{X}}$ is equipped with the $*$-weak topology $\sigma(C'_{\mathcal{X}}, C_{\mathcal{X}})$. Any Polish space $\mathcal{X}$ is equipped with its Borel $\sigma$-field and the set $\mathcal{P}_{\mathcal{X}}$ of all the probability measures on $\mathcal{X}$ is equipped with the narrow topology $\sigma(\mathcal{P}_\mathcal{X}, C_\mathcal{X})$ : the relative topology of $C'_\mathcal{X}$ on $\mathcal{P}_\mathcal{X}$. While considering random probability measures, it is necessary to equip $\mathcal{P}_\mathcal{X}$ with some $\sigma$-field: we take its Borel $\sigma$-field.

*Large deviations.* Let $\{V_n\}_{n\geq 1}$ be a sequence random variables taking their values in some topological space $\mathcal{V}$ equipped with some $\sigma$-field. One says that $\{V_n\}_{n\geq 1}$ obeys the Large Deviation Principle (LDP) in $\mathcal{V}$ with speed $n$ and rate function $I : \mathcal{V} \to [0, \infty]$, if $I$ is lower semicontinuous and for all measurable subset $A$ of $\mathcal{V}$, we have

$$
\begin{aligned}
(1.6) \qquad -\inf_{v\in \operatorname{int} A} I(y) &\leq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(V_n \in A) \\
&\leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(V_n \in A) \leq -\inf_{v\in \operatorname{cl} A} I(y)
\end{aligned}
$$

where $\operatorname{int} A$ and $\operatorname{cl} A$ are the interior and the closure of $A$ in $\mathcal{V}$.
To emphasize the parameter $n$, one says that this is a $n$-LDP. If $\rho_n$ denotes the law of $V_n$, one also writes that $\{\rho_n\}_{n\geq 1}$ obeys the $n$-LDP in $\mathcal{V}$ with the rate function $I$.
The rate function $I$ is said to be a good rate function if for each $a \geq 0$, the level set $\{I \leq a\}$ is a compact subset of $\mathcal{V}$. We shall equivalently write that $I$ is inf-compact.

1.8. **Organization of the paper.** At Section 2 the main results are stated precisely without proof. Their proofs are postponed to Section 6. They rely on preliminary results obtained at Sections 4 and 5 where general large deviation results are derived for doubly indexed sequences of random probability measures with our optimal transport problems in mind. As a preliminary approach, Section 3 is dedicated to easier analogous large deviation results in terms of simply indexed sequences. Finally, Section 7 is an appendix dedicated to the proof of a result about the Γ-convergence of convex functions which is used in Section 4. Since we didn't find this result in the literature, we give its detailed proof.

## Contents



## 2. Statement of the results

The main result of the paper is Theorem 5.1, it is stated in an abstract setting with general Polish spaces. In the present section, it is restated at Theorem 2.9 without proof in the particular framework of the optimal transport on $\mathbb{R}^d$. All the results of the present section are proved at Section 6, using the results of Sections 3, 4 and 5.

2.1. **Some transportation cost functions are LD rate functions.** Take a triangular array $(z_{n,i}\in\mathbb{R}^d; 1\le i\le n, n\ge 1)$ in $\mathbb{R}^d$ which satisfies

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \delta_{z_{n,i}}=\mu \tag{2.1}$$

for some $\mu\in\mathcal{P}_{\mathbb{R}^d}$.
For each $z\in\mathbb{R}^d$, let $\{U^k_z\}_{k\ge 1}$ be a sequence of $\mathbb{R}^d$-valued random variables. For each $k\ge 1$ and $n\ge 1$, take $n$ *independent* random variables $(X^k_{n,i}(1))_{1\le i\le n}$ where

$$X^k_{n,i}(1)\overset{Law}{=}U^k_{z_{n,i}}. \tag{2.2}$$

For each $k$, $(X^k_{n,i}(1); 1\le i\le n)_{n\ge 1}$ is a triangular array of independent particles which, in the general case, are not identically distributed because of the contribution of the deterministic $z_{n,i}$'s. An important example is given by $U^k_z=z+U^k$ with $\{U^k\}_{k\ge 1}$ a sequence of $\mathbb{R}^d$-valued random variables. This gives for each $k,n\ge 1$

$$X^k_{n,i}(1)=z_{n,i}+U^k_i \tag{2.3}$$

where $(U^k_i)_{1\le i\le n}$ are independent copies of $U^k$.

We are interested in the large deviations of the empirical measures on $\mathbb{R}^d$

$$N^k_n=\frac{1}{n}\sum_{i=1}^n \delta_{X^k_{n,i}(1)} \tag{2.4}$$

as $n$ first tends to infinity, then $k$ tends to infinity. More precisely, doubly indexed LDPs in the sense of the following definition will be proved.

**Definition 2.5** (Doubly indexed LDP)**.** Let $\mathcal{P}_{\mathcal{X}}$ be the set of all probability measures built on the Borel $\sigma$-field of a Polish space $\mathcal{X}$. The set $\mathcal{P}_{\mathcal{X}}$ is equipped with the topology of narrow convergence and with the corresponding Borel $\sigma$-filed.
One says that a doubly indexed $\mathcal{P}_{\mathcal{X}}$-valued sequence $\{L^k_n\}_{k,n\ge 1}$ obeys the $(k,n)$-LDP in $\mathcal{P}_{\mathcal{X}}$ with the rate function $I:\mathcal{P}_{\mathcal{X}}\to[0,\infty]$, if for all measurable subset $B$ of $\mathcal{P}_{\mathcal{X}}$, we have

$$\begin{aligned}
-\inf_{Q\in\operatorname{int}B} I(Q) &\le \liminf_{k\to\infty}\liminf_{n\to\infty}\frac{1}{kn}\log\mathbb{P}(L^k_n\in B)\\
&\le \limsup_{k\to\infty}\limsup_{n\to\infty}\frac{1}{kn}\log\mathbb{P}(L^k_n\in B)\le -\inf_{Q\in\operatorname{cl}B} I(Q)
\end{aligned} \tag{2.6}$$

where $\operatorname{int}B$ and $\operatorname{cl}B$ are the interior and closure of $B$ in $\mathcal{P}_{\mathcal{X}}$.

**Assumptions 2.7.** This set of assumptions holds for the present section and Section 6.

- (2.1) holds for some $\mu$ in $\mathcal{P}_{\mathbb{R}^d}$,
- for each $k\ge 1$, $(Law(U^k_z); z\in\mathbb{R}^d)$ is a Feller system in the sense of Definition 2.8 below and
- for each $z\in\mathbb{R}^d$, $\{U^k_z\}_{k\ge 1}$ obeys the $k$-LDP in $\mathbb{R}^d$ with the *good* rate function $c_z(u)\in[0,\infty]$, $u\in\mathbb{R}^d$.

**Definition 2.8.** Let $\mathcal{Z}$ and $\mathcal{X}$ be two topological spaces. The system of Borel probability measures $(P_z; z\in\mathcal{Z})$ on $\mathcal{X}$ is a *Feller system* if for all $f$ in $C_{\mathcal{X}}$, $z\in\mathcal{Z}\mapsto\int_{\mathcal{X}} f(x)\,P_z(dx)\in\mathbb{R}$ is a continuous function on $\mathcal{Z}$.

The next theorem shows that the large deviations of $\{N^k_n\}$ are closely related to optimal transport.

**Theorem 2.9.** *The doubly indexed system $\{N_n^k\}_{k,n\geq 1}$ obeys the $(k,n)$-LDP in $\mathcal{P}_{\mathbb{R}^d}$ with the rate function*

$$T(\nu) = \mathcal{T}_c(\mu,\nu)$$

*for all $\nu \in \mathcal{P}_{\mathbb{R}^d}$, where the cost function is given by*

$$c(x_0,x_1) = c_{x_0}(x_1), \quad x_0,x_1 \in \mathbb{R}^d. \tag{2.10}$$

*In the special case where (2.3) holds and $\{U^k\}_{k\geq 1}$ obeys the $k$-LDP in $\mathbb{R}^d$ with the good rate function $c : \mathbb{R}^d \to [0,\infty]$, we have $c(x_0,x_1) = c(x_1 - x_0)$, $x_0,x_1 \in \mathbb{R}^d$.*

*Examples* 2.11. In the special case where (2.3) holds, we give some examples of $\{U^k\}$ and the corresponding cost function $c$.

(1) With $U^k = Y/\sqrt{k}$ where $Y$ is a standard normal random vector on $\mathbb{R}^d$, we get

$$c(u) = |u|^2/2, \quad u \in \mathbb{R}^d.$$

This is the usual quadratic cost function.

(2) Let $(Y_m)_{m\geq 1}$ be a sequence of independent copies of a $\mathbb{R}^d$-valued random vector $Y$ which satisfies $\mathbb{E}e^{a|Y|} < \infty$ for some $a > 0$. With $U^k = \frac{1}{k}\sum_{1\leq m\leq k} Y_m$, Cramér's theorem ([8], Corollary 6.1.6) states that $\{U^k\}$ obeys the $k$-LDP in $\mathbb{R}^d$ with the rate function $c = c^Y$ :

$$c^Y(u) = \sup_{\zeta\in\mathbb{R}^d}\{\langle\zeta,u\rangle - \log \mathbb{E}e^{\langle\zeta,Y\rangle}\}, \quad u \in \mathbb{R}^d. \tag{2.12}$$

Observe that (1) is a specific instance of (2).

(3) Let $(Y_m)_{m\geq 1}$ be as above and let $\alpha$ be any continuous mapping on $\mathbb{R}^d$. With $U^k = \alpha(\frac{1}{k}\sum_{1\leq m\leq k} Y_m)$ we obtain $c(u) = \inf\{c^Y(v); v \in \mathbb{R}^d, \alpha(v) = u\}, u \in \mathbb{R}^d$ as a consequence of the contraction principle. In particular if $\alpha$ is a continuous injective mapping, then

$$c = c^Y \circ \alpha^{-1}.$$

(4) For instance, mixing (1) and (3) with $\alpha = \alpha_p$ given for each $p > 0$ and $v \in \mathbb{R}^d$ by $\alpha_p(v) = 2^{-1/p}|v|^{2/p-1}v$, taking $U^k = (2k)^{-1/p}|Y|^{2/p-1}Y$ where $Y$ is a standard normal random vector on $\mathbb{R}^d$, we get

$$c(u) = |u|^p, \quad u \in \mathbb{R}^d.$$

Note that $U^k \overset{Law}{=} k^{-1/p}Y_p$ where the density of the law of $Y_p$ is $C|z|^{p/2-1}e^{-|z|^p}$.

*Examples* 2.13. We recall some well-known examples of Cramér transform $c^Y$.

(1) To obtain the quadratic cost function $c^Y(u) = |u|^2/2$, choose $Y$ as a standard normal random vector in $\mathbb{R}^d$.
(2) Taking $Y$ such that $\mathbb{P}(Y = +1) = \mathbb{P}(Y = -1) = 1/2$, leads to
$$c^Y(u) = \begin{cases} [(1+u)\log(1+u) + (1-u)\log(1-u)]/2, & \text{if } -1 < u < +1 \\ \log 2, & \text{if } u \in \{-1,+1\} \\ +\infty, & \text{if } u \notin [-1,+1]. \end{cases}$$
(3) If $Y$ has an exponential law with expestation 1, $c^Y(u) = u - 1 - \log u$ if $u > 0$ and $c^Y(u) = +\infty$ if $u \leq 0$.
(4) If $Y$ has a Poisson law with expectation 1, $c^Y(u) = u\log u - u + 1$ if $u > 0$, $c^Y(0) = 1$ and $c^Y(u) = +\infty$ if $u < 0$.
(5) We have $c^Y(0) = 0$ if and only if $\mathbb{E}Y = 0$.
(6) More generally, $c^Y(u) \in [0,+\infty]$ and $c^Y(u) = 0$ if and only if $u = \mathbb{E}Y$.
(7) We have $c^{aY+b}(u) = c^Y((u-b)/a)$ for all real $a \neq 0$ and $b \in \mathbb{R}^d$.

*Examples* 2.14. If $\mathbb{E}Y = 0$, $c^Y$ is quadratic at the origin since $c^Y(u) = \langle u, \Gamma_Y^{-1} u\rangle/2 + o(|u|^2)$ where $\Gamma_Y$ is the covariance of $Y$. This rules out the usual costs $c(u) = |u|^p$ with $p \neq 2$.
Nevertheless, taking $Y$ a real valued variable with density $C\exp(-|z|^p/p)$ with $p \geq 1$ leads to $c^Y(u) = |u|^p/p(1 + o_{|u|\to\infty}(1))$. The case $p = 1$ follows from Example 2.13-(3) above. To see that the result still holds with $p > 1$, compute by means of the Laplace method the principal part as $\zeta$ tends to infinity of $\int_0^\infty e^{-z^p/p} e^{\zeta z}\, dz = \sqrt{2\pi(q-1)}\zeta^{1-q/2} e^{\zeta^q/q}(1 + o_{\zeta\to+\infty}(1))$ where $1/p + 1/q = 1$.
Of course, we deduce a related $d$-dimensional result considering $Y$ with the density $C\exp(-|z|_p^p/p)$ where $|z|_p^p = \sum_{i\leq d} |z_i|^p$. This gives $c^Y(u) = |u|_p^p/p(1 + o_{|u|\to\infty}(1))$.

The drawback of the specific shape of any Cramér's transform $c^Y$ (see Examples 2.14) is overcome by means of a continuous transformation as in Examples 2.11-(3 & 4).

2.2. **Convergence results.** The structure of (2.6) suggests that a $(k,n)$-LDP may be seen as the limit as $k$ tends to infinity of $n$-LDPs indexed by $k$. This is true and made precise at Proposition 2.19 and Theorem 2.20 below.
Let us have a look at the $n$-LDP satisfied by $\{N_n^k\}_{n\geq 1}$ with $k$ fixed. It is very similar to the $n$-LDP of Sanov's theorem, see Proposition 2.19 below. The only difference comes from the contribution of the initial positions $z_{n,i}$ which make $(X_{n,i}^k(1))$ a triangular array of *non*-identically independent variables. Recall that Sanov's theorem ([8], Theorem 6.2.10) states that the empirical measures $\{\frac{1}{n}\sum_{i=1}^n \delta_{X_i}\}_{n\geq 1}$ of a sequence of independent $P$-distributed random variables taking their values in a Polish space $\mathcal{X}$ obey the $n$-LDP in $\mathcal{P}_\mathcal{X}$ with the rate function

$$Q \in \mathcal{P}_\mathcal{X} \mapsto H(Q|P) = \begin{cases} \int_\mathcal{X} \log\left(\frac{dQ}{dP}\right) dQ & \text{if } Q \prec P \\ +\infty & \text{otherwise.} \end{cases} \tag{2.15}$$

$H(Q|P)$ is called the relative entropy of $Q$ with respect to $P$.
Consider now the random empirical measures on $\mathbb{R}^{2d} = \mathbb{R}^d \times \mathbb{R}^d$ which are defined by

$$M_n^k = \frac{1}{n}\sum_{i=1}^n \delta_{(z_{n,i}, X_{n,i}^k(1))} \tag{2.16}$$

for all $k, n \geq 1$. Clearly, $N_n^k$ is the second marginal of $M_n^k$. Denote for each $k \geq 1$

$$\pi^k(dx_0 dx_1) = \mu(dx_0) Law(U_{x_0}^k)(dx_1) \in \mathcal{P}_{\mathbb{R}^{2d}} \tag{2.17}$$

This means that $\pi^k = Law(X(0), X^k(1))$ where $X(0)$ is a $\mathbb{R}^d$-valued random variable which is $\mu$-distributed and $\mathbb{P}(X^k(1) \in dx_1 \mid X(0) = x_0) = Law(U_{x_0}^k)(dx_1)$. Define

$$S_k(\rho) = \begin{cases} \frac{1}{k} H(\rho|\pi^k) & \text{if } \rho_0 = \mu \\ +\infty & \text{otherwise} \end{cases}, \quad \rho \in \mathcal{P}_{\mathbb{R}^{2d}}$$

and

$$T_k(\nu) = \inf_{\rho\in\Pi(\mu,\nu)} \frac{1}{k} H(\rho|\pi^k), \quad \nu \in \mathcal{P}_{\mathbb{R}^d}. \tag{2.18}$$

**Proposition 2.19.** *For each fixed $k \geq 1$,*

(a) $\{M_n^k\}_{n\geq 1}$ *obeys the $n$-LDP in $\mathcal{P}_{\mathbb{R}^{2d}}$ with the good rate function $kS_k$ and*
(b) $\{N_n^k\}_{n\geq 1}$ *obeys the $n$-LDP in $\mathcal{P}_{\mathbb{R}^d}$ with the good rate function $kT_k$.*

The order of magnitude of $H(\rho|\pi^k)$ is $k$, since $\{U^k\}$ obeys a $k$-LDP. The rescaled entropy $S_k$ is of order 1. If it exists, $\lim_{k\to\infty} S_k$ may be interpreted as a specific entropy (see [23]). It happens that $S_k$ and $T_k$ $\Gamma$-converge. The limit of $S_k$ is

$$S(\rho) = \begin{cases} \int_{\mathbb{R}^{2d}} c\, d\rho & \text{if } \rho_0 = \mu \\ +\infty & \text{otherwise} \end{cases}, \quad \rho \in \mathcal{P}_{\mathbb{R}^{2d}}$$

where $\int_{\mathbb{R}^{2d}} c\, d\rho = \int_{\mathbb{R}^{2d}} c(x_0, x_1)\, \rho(dx_0 dx_1)$ and $c$ is given at (2.10).

**Theorem 2.20.** *We have*

(a) $\Gamma\text{-}\lim_{k\to\infty} S_k = S$ *in* $\mathcal{P}_{\mathbb{R}^{2d}}$ *and*
(b) $\Gamma\text{-}\lim_{k\to\infty} T_k = T$ *in* $\mathcal{P}_{\mathbb{R}^d}$.

These limits will allow us to deduce the following approximation results. Recall that the minimization problems ($\mathrm{MK}_k^\alpha$), ($\mathrm{MK}^\alpha$) and (MK) are defined at Section 1.3.

**Theorem 2.21.** *Assume that* $\mathcal{T}_c(\mu,\nu) < \infty$.

(a) *We have:* $\lim_{\alpha\to\infty} \lim_{k\to\infty} \inf_{\rho\in\Pi_0(\mu)} \left\{\frac{1}{k} H(\rho|\pi^k) + \alpha d(\rho_1, \nu)\right\} = \mathcal{T}_c(\mu,\nu)$.
(b) *For each $k$ and $\alpha$, ($MK_k^\alpha$) admits a unique solution $\rho_k^\alpha$ in $\mathcal{P}_{\mathbb{R}^{2d}}$. For each $\alpha$, $(\rho_k^\alpha)_{k\geq 1}$ is a relatively compact sequence in $\mathcal{P}_{\mathbb{R}^{2d}}$ and any limit point of $(\rho_k^\alpha)_{k\geq 1}$ is a solution of ($MK^\alpha$).*
(c) *For each $\alpha$, ($MK^\alpha$) admits at least a (possibly not unique) solution $\rho^\alpha$. The sequence $(\rho^\alpha)_{\alpha\geq 1}$ is relatively compact in $\mathcal{P}_{\mathbb{R}^{2d}}$ and any limit point of $(\rho^\alpha)_{\alpha\geq 1}$ is a solution of (MK).*

2.3. **The proofs.** The proofs of these announced results are done at Section 6. Theorem 2.9 is part of Theorem 6.2, Proposition 2.19 is Lemma 6.1, Theorem 2.20 is Theorem 6.5 and Theorem 2.21 is Theorem 6.6.

## 3. Large deviations of a simply indexed sequence of random measures

As a warming-up exercice, let us first consider a usual sequence of random measures.

We present an abstract setting instead of the situation described at Section 2. Let $\mathcal{X}$ and $\mathcal{Z}$ be two Polish spaces which play respectively the part of the space of "paths" $\mathbb{R}^{2d}$ and the space of initial conditions $\mathbb{R}^d$. The cost of this extension is quite low: the main property of Polish spaces to be used later is that any Borel probability measure is tight.

Take a triangular array $(z_{n,i} \in \mathcal{Z}; 1 \leq i \leq n, n \geq 1)$ on $\mathcal{Z}$ such that the sequence of empirical measures $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{z_{n,i}} \in \mathcal{P}_\mathcal{Z}$ satisfies

$$\lim_{n\to\infty} \mu_n = \mu \tag{3.1}$$

for some probability measure $\mu \in \mathcal{P}_\mathcal{Z}$. Let $(P_z \in \mathcal{P}_\mathcal{X}; z \in \mathcal{Z})$ be a collection of probability laws on $\mathcal{X}$ which is assumed to be a Feller system in the sense of Definition 2.8.
We work with a triangular array of *independent* $\mathcal{X}$-valued random variables $(X_{n,i}; 1 \leq i \leq n, n \geq 1)$ where for each index $(n,i)$ the law of $X_{n,i}$ is $P_{z_{n,i}}$. This means that for all $n \geq 1$,

$$\mathcal{L}aw(X_{n,i}; 1 \leq i \leq n) = \otimes_{1\leq i\leq n} P_{z_{n,i}}.$$

Proposition 3.14 below states a LDP in $\mathcal{P}_\mathcal{X}$ for the empirical measures

$$L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_{n,i}}$$

as $n$ tends to infinity. It is a variant of Sanov's theorem which has already been studied by Dawson and Gärtner in [7] and revisited by Cattiaux and Léonard in [6]. Nevertheless, the expression (3.15) of the rate function doesn't appear in these cited papers. The proof of Proposition 3.14 will be done as a first step for the proof of the LDP of a doubly indexed sequence: most of its ingredients will be recycled at Section 4.

**Notations.** We write shortly $\mathcal{P}_{\mathcal{Z}\mathcal{X}}$ and $C_{\mathcal{Z}\mathcal{X}}$ for $\mathcal{P}_{\mathcal{Z}\times\mathcal{X}}$ and $C_{\mathcal{Z}\times\mathcal{X}}$. The dual space $C'_{\mathcal{Z}\mathcal{X}}$ of $(C_{\mathcal{Z}\mathcal{X}}, \|\cdot\|)$ is equipped with the $*$-weak topology $\sigma(C'_{\mathcal{Z}\mathcal{X}}, C_{\mathcal{Z}\mathcal{X}})$, see Section 1.7.
Let $(Z, X)$ be the canonical projections: $Z(z,x) = z, X(z,x) = x$, $(z,x) \in \mathcal{Z}\times\mathcal{X}$. For any $q \in \mathcal{P}_{\mathcal{Z}\mathcal{X}}$, we write the desintegration

$$q(dzdx) = q_{\mathcal{Z}}(dz)q^z(dx)$$

where $q_{\mathcal{Z}}(dz) = q(Z \in dz)$ is the (marginal) law of $Z$ under $q$ and $q^z(dx) = q(X \in dx \mid Z = z), z \in \mathcal{Z}$, is a regular conditional version of the law of $X$ knowing that $Z = z$. We also define $p \in \mathcal{P}_{\mathcal{Z}\mathcal{X}}$ by

$$p(dzdx) = \mu(dz)P_z(dx).$$

The LDP for $\{L_n\}_{n\geq 1}$ will be obtained as a direct consequence of the contraction principle applied to some LDP for the sequence of $\mathcal{P}_{\mathcal{Z}\mathcal{X}}$-valued random variables

$$K_n = \frac{1}{n}\sum_{i=1}^n \delta_{(z_{n,i}, X_{n,i})}, \quad n \geq 1.$$

**Proposition 3.2.** *Suppose that (3.1) holds for some $\mu$ in $\mathcal{P}_{\mathcal{Z}}$ and that $(P_z; z \in \mathcal{Z})$ is a Feller system. Then $\{K_n\}_{n\geq 1}$ obeys the LDP in $\mathcal{P}_{\mathcal{Z}\mathcal{X}}$ with the good rate function*

$$h(q) := \begin{cases} H(q|p) = \int_{\mathcal{Z}} H(q^z|P_z)\,\mu(dz) & \text{if } q_{\mathcal{Z}} = \mu \\ +\infty & \text{otherwise} \end{cases}, \quad q \in \mathcal{P}_{\mathcal{Z}\mathcal{X}} \tag{3.3}$$

*Proof.* For all $n$ and all $F \in C_{\mathcal{Z}\mathcal{X}}$, the normalized log-Laplace transform of $K_n$ is

$$\begin{aligned} \psi_n(F) &:= \frac{1}{n}\log \mathbb{E}\exp(n\langle F, K_n\rangle) \\ &= \frac{1}{n}\sum_{i=1}^n \log \mathbb{E}e^{F(z_{n,i}, X_{n,i})} \\ &= \int_{\mathcal{Z}} \log\langle e^{F_z}, P_z\rangle\,\mu_n(dz). \end{aligned}$$

As $(\mu_n)_{n\geq 1}$ converges to $\mu$ and $(P_z; z \in \mathcal{Z})$ is a Feller system, for all $F \in C_{\mathcal{Z}\mathcal{X}}$ we have the limit:

$$\begin{aligned} \psi(F) &:= \lim_{n\to\infty} \psi_n(F) \\ &= \int_{\mathcal{Z}} \log\langle e^{F_z}, P_z\rangle\,\mu(dz). \end{aligned} \tag{3.4}$$

Following the proof of Sanov's theorem (see [8], Section 6.4) based on Dawson-Gärtner's theorem on the projective limit of LD systems (see [7], Section 3), one obtains that $\{K_n\}$ obeys the LDP in $C'_{\mathcal{Z}\mathcal{X}}$ with the rate function

$$\psi^*(q) = \sup_{F \in C_{\mathcal{Z}\mathcal{X}}} \left\{ \langle F, q\rangle - \int_{\mathcal{Z}} \log\langle e^{F_z}, P_z\rangle\,\mu(dz) \right\}, \quad q \in C'_{\mathcal{Z}\mathcal{X}}. \tag{3.5}$$

It is proved at Lemma 3.7 below, that for all $q \in C'_{\mathcal{Z}\mathcal{X}}$,

$$\psi^*(q) := \begin{cases} h(q) & \text{if } q \in \mathcal{P}_{\mathcal{Z}\mathcal{X}} \\ +\infty & \text{otherwise} \end{cases}$$

It follows that $\{K_n\}_{n\geq 1}$ obeys the LDP in $\mathcal{P}_{\mathcal{Z}\mathcal{X}}$ with the rate function $h$.

It remains to note that as the relative entropy is inf-compact and $\{q \in \mathcal{P}_{\mathcal{ZX}}; q_{\mathcal{Z}} = \mu\}$ is closed, $h$ is also inf-compact: it is a good rate function. $\square$

As a by-product of this proof, we have the following corollary which is mentioned for future use.

**Corollary 3.6.** *[Hypotheses of Proposition 3.2]. The random system $\{K_n\}_{n\geq 1}$ obeys the LDP in $C'_{\mathcal{ZX}}$ with the rate function $\psi^*$ given at (3.5).*

During the proof of Proposition 3.2 we have used the following lemma.

**Lemma 3.7.** *With $\psi^*(q)$ defined by formula (3.5) we have*

$$\psi^*(q) = \begin{cases} H(q|p) = \int_{\mathcal{Z}} H(q^z|P_z)\,\mu(dz) & \text{if } q \in \mathcal{P}_{\mathcal{ZX}} \text{ and } q_{\mathcal{Z}} = \mu \\ +\infty & \text{otherwise.} \end{cases}$$

*Proof.* The proof is twofold. We show that

(i) for all $q \in C'_{\mathcal{ZX}}$, $\psi^*(q) < +\infty$ implies that $q$ belongs to $\mathcal{P}_{\mathcal{ZX}}$ and its $z$-marginal is $q_{\mathcal{Z}} = \mu$.

(ii) Then, we show that for all $q \in \mathcal{P}_{\mathcal{ZX}}$ such that $q_{\mathcal{Z}} = \mu$, we have $\psi^*(q) = H(q|p)$.

Let $q \in C'_{\mathcal{ZX}}$ be such that

$$\sup_{F \in C_{\mathcal{ZX}}} \{\langle F, q\rangle - \psi(F)\} = \psi^*(q) < \infty.$$

Let us begin with the proof of (i).

• Let us show that $q \geq 0$. Let $F_o \in C_{\mathcal{ZX}}$ be such that $F_o \geq 0$. As $\psi(aF_o) \leq 0$ for all $a \leq 0$,

$$\begin{aligned} \psi^*(q) &\geq \sup_{a\leq 0}\{a\langle F_o, q\rangle - \psi(aF_o)\} \\ &\geq \sup_{a\leq 0}\{a\langle F_o, q\rangle\} \\ &= \begin{cases} 0, & \text{if } \langle F_o, q\rangle \geq 0 \\ +\infty, & \text{otherwise} \end{cases} \end{aligned}$$

Therefore, as $\psi^*(q) < \infty$, $\langle F_o, q\rangle \geq 0$ for all $F_o \geq 0$, which is the desired result.

• Let us show that $\langle \mathbf{1}, q\rangle = 1$. For any constant function $F \equiv c \in \mathbb{R}$, we have $\psi(c\mathbf{1}) = c$. It follows that

$$\begin{aligned} \psi^*(q) &\geq \sup_{c\in\mathbb{R}}\{c\langle \mathbf{1}, q\rangle - \psi(c\mathbf{1})\} \\ &\geq \sup_{c\in\mathbb{R}}\{c(\langle \mathbf{1}, q\rangle - 1)\} \\ &= \begin{cases} 0, & \text{if } \langle \mathbf{1}, q\rangle = 1 \\ +\infty, & \text{otherwise} \end{cases} \end{aligned}$$

from which the result follows.

• In order to prove that $q$ is $\sigma$-additive, we have to prove that for any sequence $(F_n)_{n\geq 1}$ in $C_{\mathcal{ZX}}$ such that $F_n \geq 0$ for all $n$ and $(F_n(z,x))_{n\geq 1}$ decreases to zero for each $(z,x) \in \mathcal{Z}\times\mathcal{X}$, we have

$$\lim_{n\to\infty} \langle F_n, q\rangle = 0. \tag{3.8}$$

For such a sequence, one can apply the dominated convergence theorem to obtain that

$$\lim_{n\to\infty} \psi(aF_n) = 0,$$

for all $a \geq 0$. It follows that for all $q \in C'_{\mathcal{ZX}}$,

$$\begin{aligned} \psi^*(q) &\geq \sup_{a\geq 0} \limsup_{n\to\infty} \{a\langle F_n, q\rangle - \psi(aF_n)\} \\ &\geq \sup_{a\geq 0} \left( \limsup_{n\to\infty} a\langle F_n, q\rangle - \lim_{n\to\infty} \psi(aF_n) \right) \\ &= \sup_{a\geq 0} a \limsup_{n\to\infty} \langle F_n, q\rangle \\ &= \begin{cases} 0 & \text{if } \limsup_{n\to\infty}\langle F_n, q\rangle \leq 0 \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Therefore, as $\psi^*(q) < \infty$, we have $\limsup_{n\to\infty}\langle F_n, q\rangle \leq 0$. Since we have just seen that $q \geq 0$, we have obtained (3.8).

This completes the proof of $q \in \mathcal{P}_{\mathcal{ZX}}$ since we have proved that any $q \in C'_{\mathcal{ZX}}$ such that $\psi^*(q) < \infty$ is nonnegative, has a unit mass and satisfies (3.8). Therefore, $q$ is uniquely identified with a probability measure on the Polish space $\mathcal{Z} \times \mathcal{X}$ (see [18], Proposition II-7-2).

• To complete the proof of (i), it remains to show that for any $q \in \mathcal{P}_{\mathcal{ZX}}$, $\psi^*(q) < \infty$ implies that $q_{\mathcal{Z}} = \mu$. Indeed, choosing $F(z, x) = g(z)$ not depending on $x$ with $g \in C_{\mathcal{Z}}$, one sees that

$$\begin{aligned} \psi^*(q) &\geq \sup_{g\in C_{\mathcal{Z}}} \{\langle g, q_{\mathcal{Z}}\rangle - \langle g, \mu\rangle\} \\ &= \begin{cases} 0, & \text{if } q_{\mathcal{Z}} = \mu \\ +\infty, & \text{otherwise} \end{cases} \end{aligned}$$

which gives the announced result.

Now, let us show (ii). For all $q \in \mathcal{P}_{\mathcal{ZX}}$ such that $q_{\mathcal{Z}} = \mu$ or equivalently such that $q(dzdx) = \mu(dz)q^z(dx)$, we have

$$\begin{aligned} \psi^*(q) &= \sup_{F\in C_{\mathcal{ZX}}} \int_{\mathcal{Z}} (\langle F_z, q^z\rangle - \log\langle e^{F_z}, P_z\rangle)\, \mu(dz) \\ &\leq \int_{\mathcal{Z}} \sup_{f\in C_{\mathcal{X}}} \{\langle f, q^z\rangle - \log\langle e^f, P_z\rangle\}\, \mu(dz) \\ &\overset{(a)}{=} \int_{\mathcal{Z}} H(q^z|P_z)\, \mu(dz) \\ &\overset{(b)}{=} H(q|p) \end{aligned} \tag{3.9}$$

where equality (a) follows from the well-known variational representation of the relative entropy in a Polish space $\mathcal{X}$

$$H(Q|P) = \sup_{f\in C_{\mathcal{X}}} \left\{ \int_{\mathcal{X}} f\, dQ - \log \int_{\mathcal{X}} e^f\, dP \right\}, \quad P, Q \in \mathcal{P}_{\mathcal{X}} \tag{3.10}$$

and equality (b) follows from the tensorization property

$$H(q|p) = H(q_{\mathcal{Z}}|p_{\mathcal{Z}}) + \int_{\mathcal{Z}} H(q^z|p^z)\, q_{\mathcal{Z}}(dz) \tag{3.11}$$

since $p_{\mathcal{Z}} = q_{\mathcal{Z}} = \mu$ and $p^z = p(\cdot \mid Z = z) = P_z$. Note that $z \mapsto H(q^z|P_z)$ is measurable. Indeed, $(Q, P) \mapsto H(Q|P)$ is measurable as a lower semicontinuous function and $z \mapsto (q^z, P_z)$ is measurable since its coordinates are measurable: $z \mapsto q^z$ is measurable as a regular conditional version in a Polish space and $z \mapsto P_z$ is assumed to be continuous. We have just proved that $\psi^* \le h$.

The converse inequality follows from Jensen's inequality: $\psi(F) \le \log \int_{\mathcal{Z}} \langle e^{F_z}, P_z\rangle \, \mu(dz) = \log\langle e^F, p\rangle$ for all $F \in C_{\mathcal{Z}\mathcal{X}}$. Indeed, taking the convex conjugates leads us for all $q \in \mathcal{P}_{\mathcal{Z}\mathcal{X}}$ to

$$\begin{aligned} \psi^*(q) &\ge \sup_{F \in C_{\mathcal{Z}\mathcal{X}}} \{\langle F, q\rangle - \log\langle e^F, p\rangle\} \\ &= H(q|p) \end{aligned} \tag{3.12}$$

This equality is (3.10). This completes the proof of the lemma. □

*Remark* 3.13. The $\|\cdot\|$-continuity of $q$ in $C'_{\mathcal{Z}\mathcal{X}}$ didn't play any role in the proof. Only its linearity has been used.

Now, we investigate the large deviations of

$$L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_{n,i}}.$$

Let us denote the $\mathcal{X}$-marginal of $p$ by

$$P(dx) = \int_{\mathcal{Z}} P_z(dx)\, \mu(dz) \in \mathcal{P}_{\mathcal{X}}.$$

**Proposition 3.14.** *Suppose that (3.1) holds for some $\mu$ in $\mathcal{P}_{\mathcal{Z}}$ and that $(P_z; z \in \mathcal{Z})$ is a Feller system.*

(a) *Then, $\{L_n\}_{n\ge 1}$ obeys the LDP in $\mathcal{P}_{\mathcal{X}}$ with the good rate function $H$ which is defined for all $Q \in \mathcal{P}_{\mathcal{X}}$ by*

$$H(Q) = \inf\left\{\int_{\mathcal{Z}} H(\Pi_z|P_z)\,\mu(dz); (\Pi_z)_{z\in\mathcal{Z}} : \int_{\mathcal{Z}} \Pi_z\,\mu(dz) = Q\right\} \tag{3.15}$$

*where the transition kernels $z \in \mathcal{Z} \mapsto \Pi_z \in \mathcal{P}_{\mathcal{X}}$ are measurable.*

(b) *If $H(Q) < +\infty$, there exists a unique (up to $\mu$-a.e. equality) kernel $(\Pi^*_z)_{z\in\mathcal{Z}}$ which realizes the infimum in (3.15): $H(Q) = \int_{\mathcal{Z}} H(\Pi^*_z|P_z)\,\mu(dz)$.*

(c) *If in addition the Feller system $(P_z)_{z\in\mathcal{Z}}$ satisfies*

$$P_z = P(\cdot \mid \beta(X) = z)$$

*for $\mu$-almost every $z \in \mathcal{Z}$ and some continuous function $\beta : \mathcal{X} \to \mathcal{Z}$, we have for all $Q \in \mathcal{P}_{\mathcal{X}}$*

$$H(Q) = \begin{cases} H(Q|P) & \text{if } Q \text{ satifies } \beta \diamond Q = \mu \\ +\infty & \text{otherwise.} \end{cases}$$

*and the minimizing kernel $(\Pi^*_z)$ of (3.15) is $\Pi^*_z = Q(\cdot \mid \beta(X) = z)$, for $\mu$-almost every $z$.*

*Proof.* Let us prove (a). As $L_n$ is the $\mathcal{X}$-marginal of $K_n$ and $\{K_n\}$ obeys the LDP with a good rate function, the statement (a) follows from the contraction principle (see [8], Theorem 4.2.1): $\{L_n\}$ obeys the LDP in $\mathcal{P}_{\mathcal{X}}$ with the good rate function $H(Q) = \inf\{h(q); q \in \mathcal{P}_{\mathcal{Z}\mathcal{X}} : q_{\mathcal{X}} = Q\}$ which is (3.15).

The statement (b) immediately follows from the strict convexity and the inf-compactness of $q \mapsto H(q|p)$ which is restricted to the closed convex set $\{q \in \mathcal{P}_{\mathcal{Z}\mathcal{X}} : q_{\mathcal{Z}} = \mu, q_{\mathcal{X}} = Q\}$.

Let us prove (c). To do this, we rewrite the proof of Proposition 3.2 with $L_n$ instead of $K_n$. We obtain that $\{L_n\}$ obeys the LDP in $\mathcal{P}_\mathcal{X}$ with the rate function

$$\Psi^*(Q) = \sup_{f\in C_\mathcal{X}} \left\{ \langle f, Q\rangle - \int_\mathcal{Z} \log\langle e^f, P_z\rangle\, \mu(dz) \right\} \tag{3.16}$$

This equality is (3.5) where we replace $q$ by $Q$ and $F(z,x)$ by $f(x)$. Choosing $f \in C_\mathcal{X}$ of the form $f = g\circ\beta$ with $g$ in $C_\mathcal{Z}$ gives us for all $Q \in \mathcal{P}_\mathcal{X}$

$$\begin{aligned} \Psi^*(Q) &\geq \sup_{g\in C_\mathcal{Z}} \{\langle g, \beta\diamond Q\rangle - \langle g, \mu\rangle\} \\ &= \begin{cases} 0 & \text{if } \beta\diamond Q = \mu \\ +\infty & \text{otherwise} \end{cases} \end{aligned}$$

It follows that $\Psi^*(Q) < \infty$ implies that $\beta\diamond Q = \mu$. For such a $Q$, as in the proof of inequality (3.9), we obtain the inequality in $\Psi^*(Q) \leq \int_\mathcal{Z} H\big(Q(\cdot|Z = z) \mid P_z\big)\, \mu(dz) = H(Q|P)$. This last equality follows from the tensorization property of the relative entropy, see (3.11). This proves that $\Psi^* \leq H$. The converse inequality follows from Jensen's inequality exactly as in the proof of inequality (3.12). We have shown that

$$\Psi^* = H. \tag{3.17}$$

The last statement about the minimizing kernel is a direct consequence of the tensorization formula (3.11):

$$\begin{aligned} &\inf\left\{ \int_\mathcal{Z} H(\Pi_z|P_z)\,\mu(dz); (\Pi_z)_{z\in\mathcal{Z}} : \int_\mathcal{Z} \Pi_z\,\mu(dz) = Q \right\} \\ =\ & H(Q|P) \\ =\ & H(Q_\mathcal{Z}|P_\mathcal{Z}) + \int_\mathcal{Z} H\big(Q(\cdot|Z = z) \mid P(\cdot|Z = z)\big)\, Q_\mathcal{Z}(dz) \\ =\ & \int_\mathcal{Z} H\big(Q(\cdot|Z = z) \mid P_z\big)\, \mu(dz) \end{aligned}$$

where the first equality follows from (a) and the first part of this statement, and the last equality follows from $H(Q_\mathcal{Z}|P_\mathcal{Z}) = H(\mu|\mu) = 0$. □

*Remark* 3.18. The identity (3.15) is a formal inf-convolution formula and (3.16) is its dual formulation: "the convex conjugate of an inf-convolution is the sum of the convex conjugates".

*Remark* 3.19. Statement (c) holds true also when $\beta$ is only assumed to be measurable. Indeed, (3.16) can be strengthened by

$$\Psi^*(Q) = \sup_{f\in C_\mathcal{X}} \left\{ \langle f, Q\rangle - \int_\mathcal{Z} \log\langle e^f, P_z\rangle\, \mu(dz) \right\} = \sup_{f\in B(\mathcal{X})} \left\{ \langle f, Q\rangle - \int_\mathcal{Z} \log\langle e^f, P_z\rangle\, \mu(dz) \right\},$$

for all $Q \in \mathcal{P}_\mathcal{X}$, where $B(\mathcal{X})$ is the space of all measurable bounded functions on $\mathcal{X}$. For the second equality, note that in the proof of Proposition 3.2, taking the test functions $F(z,x)$ bounded, $z$-continuous and $x$-measurable (instead of $x$-continuous), does not change anything except that in the expression of the rate function $\sup_F\{\langle F, q\rangle - \psi(F)\}$, the sup is taken over this larger space instead of $C_{\mathcal{Z}\mathcal{X}}$. As the rate function is unique, the sup over these two spaces is the same. A similar argument in the present situation leads to $\sup_{f\in C_\mathcal{X}} = \sup_{f\in B(\mathcal{X})}$. Finally, choosing $f \in B(\mathcal{X})$ of the form $f = g\circ\beta$ with $g$ in $C_\mathcal{Z}$ gives us $\Psi^*(Q) \geq \sup_{g\in C_\mathcal{Z}}\{\langle g, \beta\diamond Q\rangle - \langle g, \mu\rangle\}$ and one concludes as in the previous proof.

## 4. Large deviations of a doubly indexed sequence of random measures. Preliminary results

We keep the abstract Polish spaces $\mathcal{Z}$ and $\mathcal{X}$ of Section 3, as well as the triangular array $(z_{n,i} \in \mathcal{Z}; 1 \leq i \leq n, n \geq 1)$ which satisfies (3.1). For each $k \geq 1$, we consider a Feller system $(P^k_z \in \mathcal{P}_\mathcal{X}; z \in \mathcal{Z})$ of probability laws on $\mathcal{X}$ and a triangular array of *independent* $\mathcal{X}$-valued random variables $(X^k_{n,i}; 1 \leq i \leq n, n \geq 1)$ where for each index $(n,i)$ the law of $X^k_{n,i}$ is $P^k_{z_{n,i}}$. This means that for all $k, n \geq 1$,

$$\mathcal{L}aw(X^k_{n,i}; 1 \leq i \leq n) = \otimes_{1\leq i\leq n} P^k_{z_{n,i}}.$$

The main result of the next Section 5 states the $(k,n)$-LDP in $\mathcal{P}_\mathcal{X}$ for

$$L^k_n = \frac{1}{n}\sum_{i=1}^n \delta_{X^k_{n,i}}, \quad k, n \geq 1.$$

As in Section 3, this LDP will be obtained by means of the contraction principle applied to some LDP for the $\mathcal{P}_{\mathcal{Z}\mathcal{X}}$-valued random variables

$$K^k_n = \frac{1}{n}\sum_{i=1}^n \delta_{(z_{n,i}, X^k_{n,i})}, \quad k, n \geq 1.$$

The main result of the present section is Theorem 4.9. It states the $(k,n)$-LDP for $\{K^k_n\}_{k,n\geq 1}$.

We also assume that for each $z \in \mathcal{Z}$, $(P^k_z)_{k\geq 1}$ obeys some $k$-LDP in $\mathcal{X}$ with rate function $J_z$. This means that for each $k$ and all measurable subset $A$ of $\mathcal{X}$

$$\begin{aligned}
-\inf_{x\in \mathrm{int}\, A} J_z(x) &\leq \liminf_{k\to\infty} \frac{1}{k}\log P^k_z(A) \\
&\leq \limsup_{k\to\infty} \frac{1}{k}\log P^k_z(A) \leq -\inf_{x\in \mathrm{cl}\, A} J_z(x)
\end{aligned}$$

where $\mathrm{int}\, A$ and $\mathrm{cl}\, A$ are the interior and the closure of $A$ in $\mathcal{X}$. This plays the part of Cramér's theorem and its transformations at Section 2, see Examples 2.11 with $J_z(x) = c^Y(x_1 - x_0)$, $x = (x_0, x_1) \in \mathbb{R}^{2d}$, $z \in \mathbb{R}^d$ if $x_0 = z$ and $+\infty$ otherwise.

### 4.1. Preliminary results.

Before proving the $(k,n)$-LDP for $\{K^k_n\}$ at Theorem 4.9, we need some preliminary results. The following lemma is Corollary 7.4, its detailed proof is given at Section 7.

**Lemma 4.1.** *Let $(\mathcal{F}, \|\cdot\|)$ be a normed vector space and $\mathcal{Q}$ be its dual space. Let $\lambda$, $\lambda_k$, $k \geq 1$ be real-valued convex functions on $\mathcal{F}$ such that*

(a) $\lim_{k\to\infty} \lambda_k(F) = \lambda(F)$ *for all $F \in \mathcal{F}$ and*
(b) *there exists $c > 0$ such that* $\sup_{k\geq 1} |\lambda_k(F)| \leq c(1 + \|F\|)$ *for all $F \in \mathcal{F}$.*

*Then, the convex conjugates $\lambda^*_k$ of $\lambda_k$, $\Gamma$-converge to the convex conjugate $\lambda^*$ of $\lambda$ :*

$$\Gamma\text{-}\lim_{k\to\infty} \lambda^*_k(q) = \lambda^*(q)$$

*for all $q \in \mathcal{Q}$, with respect to the $*$-weak topology $\sigma(\mathcal{Q}, \mathcal{F})$.*

The following lemma is proved in [14].

**Lemma 4.2.** *Suppose that for all $k \geq 1$, $\{\mu^k_n\}_{n\geq 1}$ obeys a weak $n$-LDP with rate function $kI^k$ and also suppose that the sequence $(I^k)_{k\geq 1}$ $\Gamma$-converges to some function $I$. Then, $\{\mu^k_n\}_{k,n\geq 1}$ obeys a weak $(k,n)$-LDP with rate function $I$.*

*Proof.* See [14]. □

We define for each $k$ and all $F \in C_{\mathcal{Z}\mathcal{X}}$,

$$
\begin{aligned}
\lambda_k(F) &= \frac{1}{k}\int_{\mathcal{Z}} \log\langle e^{kF_z}, P^k_z\rangle\, \mu(dz) \\
\lambda(F) &= \int_{\mathcal{Z}} \sup_{x\in\mathcal{X}}\{F(z,x) - J_z(x)\}\, \mu(dz).
\end{aligned} \tag{4.3}
$$

Note that $z \mapsto \sup_{x\in\mathcal{X}}\{F(z,x) - J_z(x)\}$ is measurable since it is the pointwise limit of continuous functions: see (4.5) below, so that $\lambda(F)$ is well-defined.
Observe that $\lambda_k$ is a normalized version of the function $\psi$ defined at (3.4).

**Lemma 4.4.** *We assume that for each $z \in \mathcal{Z}$, $(P^k_z)_{k\geq 1}$ obeys the $k$-LDP in $\mathcal{X}$ with the* good *rate function $J_z$. Then, for all $F \in C_{\mathcal{Z}\mathcal{X}}$ we have*

$$\lim_{k\to\infty} \frac{1}{k}\log\langle e^{kF_z}, P^k_z\rangle = \sup_{x\in\mathcal{X}}\{F(z,x) - J_z(x)\}, \tag{4.5}$$

$$\lim_{k\to\infty} \lambda_k(F) = \lambda(F) \quad \textit{and} \tag{4.6}$$

$$\sup_k |\lambda_k(F)| \leq \|F\|, \quad |\lambda(F)| \leq \|F\|, \quad \forall F \in C_{\mathcal{Z}\mathcal{X}}. \tag{4.7}$$

*The functions $\lambda_k$ and $\lambda$ are convex and $\sigma(C_{\mathcal{Z}\mathcal{X}}, C'_{\mathcal{Z}\mathcal{X}})$-lower semicontinuous..*

*Proof.* Thanks to the assumption on $(P^k_z)_{k\geq 1}$, by Varadhan's integral lemma (see [8], Theorem 4.3.1), as $F_z$ is continuous and bounded and $J_z$ is assumed to be a good rate function, for all $z$ we have (4.5).

As for all $k \geq 1$, $z \in \mathcal{Z}$ and $F \in C_{\mathcal{Z}\mathcal{X}}$, we have $|\frac{1}{k}\log\langle e^{kF_z}, P^k_z\rangle| \leq \|F\|$, with (4.5) we see that

$$\left|\sup_{x\in\mathcal{X}}\{F(z,x) - J_z(x)\}\right| \leq \|F\|. \tag{4.8}$$

These estimates allow us to apply Lebesgue dominated convergence theorem to obtain (4.6) and (4.7).

For each $k$, $\lambda_k$ is convex since $f \mapsto \log\langle e^{kf}, P^k_z\rangle$ is convex as a log-Laplace transform and $\mu$ is a nonnegative measure. As a pointwise limit of convex functions, $\lambda$ is also convex.

The convex functions $\lambda^k$ and $\lambda$ are $\sigma(C_{\mathcal{Z}\mathcal{X}}, C'_{\mathcal{Z}\mathcal{X}})$-lower semicontinuous if and only if they are $\|\cdot\|$-lower semicontinuous on $C_{\mathcal{Z}\mathcal{X}}$. But, because of (4.7), these convex functions are $\|\cdot\|$-continuous on the whole space $C_{\mathcal{Z}\mathcal{X}}$. A fortiori, they are lower semicontinuous. □

4.2. **The $(k,n)$-LDP for $\{K^k_n\}$.** Let us introduce the convex conjugate of $\lambda$ :

$$\lambda^*(q) = \sup_{F\in C_{\mathcal{Z}\mathcal{X}}} \left\{\langle q, F\rangle - \int_{\mathcal{Z}} \sup_{x\in\mathcal{X}}\{F(z,x) - J_z(x)\}\, \mu(dz)\right\}, \quad q \in C'_{\mathcal{Z}\mathcal{X}}.$$

It will appear during the proof of Theorem 4.9 that it is the rate function of the $(k,n)$-LDP satisfied by $\{K^k_n\}_{k,n\geq 1}$.

**Theorem 4.9.** *Suppose that*

1. *$(\mu_n)_{n\geq 1}$ converges to $\mu$ in $\mathcal{P}_{\mathcal{Z}}$,*
2. *for each $k \geq 1$, $(P^k_z; z \in \mathcal{Z})$ is a Feller system in the sense of Definition 2.8,*
3. *for each $z \in \mathcal{Z}$, $(P^k_z)_{k\geq 1}$ obeys the $k$-LDP in $\mathcal{X}$ with the good rate function $J_z$.*

*Then* $\{K_n^k\}_{k,n\geq 1}$ *obeys the* $(k,n)$*-LDP in* $\mathcal{P}_{\mathcal{ZX}}$ *with the affine good rate function*

$$i(q) := \begin{cases} \int_{\mathcal{Z}\times\mathcal{X}} J_z(x)\, q(dzdx) & \text{if } q_{\mathcal{Z}} = \mu \\ +\infty & \text{otherwise} \end{cases}, \quad q \in \mathcal{P}_{\mathcal{ZX}}. \tag{4.10}$$

*Proof.* The framework of the proof is the same as Proposition 3.2's one, but it is technically more demanding.
For all $k, n \geq 1$ and all $F \in C_{\mathcal{ZX}}$, the normalized log-Laplace transform of $K_n^k$ is defined by

$$\lambda_k^n(F) := \frac{1}{kn} \log \mathbb{E} \exp(kn\langle F, K_n^k\rangle) = \frac{1}{k} \int_{\mathcal{Z}} \log\langle e^{kF_z}, P_z^k\rangle\, \mu_n(dz).$$

For fixed $k$, considering the limit as $n$ tends to infinity and taking assumptions (1) and (2) into account gives

$$\lim_{n\to\infty} \lambda_k^n(F) = \lambda_k(F).$$

By Corollary 3.6 we see that for all $k$, $\{K_n^k\}_{n\geq 1}$ obeys the $n$-LDP in $C'_{\mathcal{ZX}}$ with the rate function $\lambda_k^*/k$.

Because of Lemma 4.4 and Lemma 4.1 applied with $\mathcal{F} = C_{\mathcal{ZX}}$ and $\mathcal{Q} = C'_{\mathcal{ZX}}$, the pointwise convergence (4.6) and the estimate (4.7) imply that

$$\Gamma\text{-}\lim_{k\to\infty} \lambda_k^* = \lambda^* \tag{4.11}$$

in $C'_{\mathcal{ZX}}$.
By Lemma 4.2, this $\Gamma$-convergence implies that $\{K_n^k\}_{k,n\geq 1}$ obeys a *weak* $(k,n)$-LDP in $C'_{\mathcal{ZX}}$ with the rate function $\lambda^*$. It is proved at Lemma 4.13 below that

$$\{\lambda^* < +\infty\} \subset \mathcal{P}_{\mathcal{ZX}}. \tag{4.12}$$

A fortiori, $\{\lambda^* < +\infty\}$ is included in the strong unit ball

$$U_{\mathcal{ZX}} = \left\{ q \in C'_{\mathcal{ZX}};\, \|q\|^* := \sup_{F\in C_{\mathcal{ZX}}, \|F\|\leq 1} \langle q, F\rangle \leq 1 \right\}$$

of $C'_{\mathcal{ZX}}$ which is $\sigma(C'_{\mathcal{ZX}}, C_{\mathcal{ZX}})$-compact (Banach-Alaoglu theorem). Consequently, $\{K_n^k\}_{k,n\geq 1}$ obeys a *strong* $(k,n)$-LDP in $U_{\mathcal{ZX}}$ with the topology $\sigma(U_{\mathcal{ZX}}, C_{\mathcal{ZX}})$ and the rate function $\lambda^*$. With (4.12) again, we obtain that $\{K_n^k\}_{k,n\geq 1}$ obeys the $(k,n)$-LDP in $\mathcal{P}_{\mathcal{ZX}}$ with the rate function $\lambda^*$.

Let us show that the restriction of $\lambda^*$ to $\mathcal{P}_{\mathcal{ZX}}$ has $\sigma(\mathcal{P}_{\mathcal{ZX}}, C_{\mathcal{ZX}})$-compact level sets. As a convex conjugate, $\lambda^*$ is $\sigma(C'_{\mathcal{ZX}}, C_{\mathcal{ZX}})$-lower semicontinuous. Therefore, for all real $a$, $\{\lambda^* \leq a\}$ is $\sigma(C'_{\mathcal{ZX}}, C_{\mathcal{ZX}})$-closed. But, (4.12) implies that $\{\lambda^* \leq a\}$ is included in the $\sigma(C'_{\mathcal{ZX}}, C_{\mathcal{ZX}})$-compact unit ball $U_{\mathcal{ZX}}$. Hence, $\{\lambda^* \leq a\}$ is $\sigma(C'_{\mathcal{ZX}}, C_{\mathcal{ZX}})$-compact and by (4.12) again, the restriction of $\lambda^*$ to $\mathcal{P}_{\mathcal{ZX}}$ is $\sigma(\mathcal{P}_{\mathcal{ZX}}, C_{\mathcal{ZX}})$-inf-compact.

Finally, it will be proved at Proposition 4.14 that the restriction of $\lambda^*$ to $\mathcal{P}_{\mathcal{ZX}}$ is $i$. This completes the proof of the theorem. ☐

4.3. **Identification of the rate function** $\lambda^*$**.** It remains to show that $\lambda^* = i$. This is the most technical part of the paper.

**Lemma 4.13.** *Under the assumptions of Lemma 4.4, the following statements hold true.*

(a) *For all* $q \in C'_{\mathcal{ZX}}$, $\lambda^*(q) < \infty$ *implies that* $q \in \mathcal{P}_{\mathcal{ZX}}$.
(b) *For all* $q \in \mathcal{P}_{\mathcal{ZX}}$, $\lambda^*(q) < \infty$ *implies that* $q_{\mathcal{Z}} = \mu$.

*Proof.* It is similar to the proof of Lemma 3.7. As in Lemma 3.7, the $\|\cdot\|$-continuity of $q$ doesn't play any role, see Remark 3.13. Let $q \in C'_{\mathcal{Z}\mathcal{X}}$ be such that

$$\sup_{F\in C_{\mathcal{Z}\mathcal{X}}} \{\langle F, q\rangle - \lambda(F)\} = \lambda^*(q) < \infty.$$

An inspection of Lemma 3.7's proof shows that, to prove that $q \in \mathcal{P}_{\mathcal{Z}\mathcal{X}}$, it is enough to check that $\lambda$ satisfies

(i) for all $a \le 0$ and all nonnegative $F_o \in C_{\mathcal{Z}\mathcal{X}}$, $\lambda(aF_o) \le 0$
(ii) for any constant function $F \equiv c \in \mathbb{R}$, $\lambda(c\mathbf{1}) = c$
(iii) for any sequence $(F_n)_{n\ge1}$ in $C_{\mathcal{Z}\mathcal{X}}$ such that $F_n \ge 0$ for all $n$ and $(F_n(z,x))_{n\ge1}$ decreases to zero for each $(z,x) \in \mathcal{Z}\times\mathcal{X}$, we have, $\lim_{n\to\infty}\lambda(aF_n) = 0$, for all $a \ge 0$.

(i) As $J_z(x) \ge 0$ for all $z$ and $x$, and $\mu \ge 0$, we have $\lambda(aF_o) \le 0$ for all $a \le 0$ and all nonnegative $F_o \in C_{\mathcal{Z}\mathcal{X}}$.

(ii) As $\inf_{x\in\mathcal{X}} J_z(x) = 0$ for all $z$ and $\mu$ is a probability measure, for any constant function $F \equiv c \in \mathbb{R}$, we have $\lambda(c\mathbf{1}) = c$.

(iii) By Lemma 4.26 below, for all $z \in \mathcal{Z}$, $\big(\sup_{x\in\mathcal{X}}\{F_n(z,x) - J_z(x)\}\big)_{n\ge1}$ is a decreasing sequence and $\lim_{n\to\infty}\sup_{x\in\mathcal{X}}\{F_n(z,x) - J_z(x)\} = 0$. As $|\sup_{x\in\mathcal{X}}\{F_n(z,x) - J_z(x)\}| \le \sup_{z,x}|F_1(z,x)| < \infty$ for all $n$ and $z$, one can apply the dominated convergence theorem to obtain that $\lim_{n\to\infty}\lambda(aF_n) = 0$, for all $a \ge 0$.

This completes the proof of statement (a).

Let us prove (b). Choosing $F(z,x) = g(z)$ not depending on $x$ with $g \in C_{\mathcal{Z}}$ in the expression of $\lambda^*$, one sees that for all $q \in \mathcal{P}_{\mathcal{Z}\mathcal{X}}$

$$\begin{aligned}\lambda^*(q) &\ge \sup_{g\in C_{\mathcal{Z}}}\{\langle g, q_{\mathcal{Z}}\rangle - \langle g, \mu\rangle\}\\ &= \begin{cases} 0, & \text{if } q_{\mathcal{Z}} = \mu\\ +\infty, & \text{otherwise}\end{cases}\end{aligned}$$

which gives the announced result and completes the proof of Lemma 4.13. □

The very technical result of this section is the following Proposition 4.14. During its proof, we need some lemmas whose statements are included in the body of the proof. The proofs of these lemmas are postponed to the next subsection 4.4.

**Proposition 4.14.** *For all $q \in \mathcal{P}_{\mathcal{Z}\mathcal{X}}$, $\lambda^*(q) = i(q)$.*

*Proof.* Thanks to Lemma 4.13-b, to prove that $\lambda^* = i$, we have to show that $\lambda^*(q) = \int_{\mathcal{Z}\times\mathcal{X}} J_z(x)\, q(dzdx)$ for all $q \in \mathcal{P}_{\mathcal{Z}\mathcal{X}}$ such that $q_{\mathcal{Z}} = \mu$ or equivalently such that

$$q(dzdx) = \mu(dz)q^z(dx) \tag{4.15}$$

where

$$q^z(dx) = q(X \in dx \mid Z = z).$$

For such a $q$ we have

$$\begin{aligned}
\lambda^*(q) &= \sup_{F\in C_{\mathcal{ZX}}} \int_{\mathcal{Z}} [\langle F_z, q^z\rangle - \sup_{x\in\mathcal{X}}\{F_z(x) - J_z(x)\}]\,\mu(dz)\\
&\leq \int_{\mathcal{Z}} \sup_{f\in C_{\mathcal{X}}} [\langle f, q^z\rangle - \sup_{x\in\mathcal{X}}\{f(x) - J_z(x)\}]\,\mu(dz)\\
&\stackrel{(a)}{=} \int_{\mathcal{Z}\times\mathcal{X}} J_z(x)\, q^z(dx)\mu(dz)\\
&\stackrel{(b)}{=} i(q)
\end{aligned}$$

where equality (a) is given at the following Lemma 4.16 and equality (b) follows from (4.15).

**Lemma 4.16.** *Let $J$ be a $[0, +\infty]$-valued lower semicontinuous function on $\mathcal{X}$. For all $Q \in \mathcal{P}_{\mathcal{X}}$, we have*

$$\sup_{f\in C_{\mathcal{X}}} \left\{ \int_{\mathcal{X}} f\, dQ - \sup_{x\in\mathcal{X}}(f(x) - J(x)) \right\} = \int_{\mathcal{X}} J\, dQ.$$

The proof of this lemma is put back after the proof of the present proposition.

Note that $z \mapsto \langle J_z, q^z\rangle$ is measurable since $z \mapsto J_z(x)$ is assumed to be continuous for all $x$ and $z \mapsto q^z$ is a regular version of the desintegration of $q$.

It remains to show the converse inequality: $\lambda^*(q) \geq i(q)$ for all $q$ satisfying (4.15). As a first step, we would like to invert a sup and an integral to obtain

$$\begin{aligned}
\lambda^*(q) &= \sup_{F\in C_{\mathcal{ZX}})} \int_{\mathcal{Z}} [\langle F_z, q^z\rangle - \sup_{x\in\mathcal{X}}\{F_z(x) - J_z(x)\}]\,\mu(dz)\\
&= \int_{\mathcal{Z}} \sup_{f\in C_{\mathcal{X}}} [\langle f, q^z\rangle - \sup_{x\in\mathcal{X}}\{f(x) - J_z(x)\}]\,\mu(dz) \qquad (4.17)
\end{aligned}$$

As a first step, we are going to prove this equality under the restrictive assumption that $\mathcal{X}$ is compact. Its proof relies on the following result which is due to R. T. Rockafellar (see [20], Theorem 2).

**Lemma 4.18.** *Let $(\mathcal{Z}, \mu)$ be a measure space such that $\mu$ is $\sigma$-finite. Let $L$ be a* decomposable space *(see below for the definition) of measurable functions $F$ on $\mathcal{Z}$ with their values in a Polish space $\mathcal{Y}$ equipped with its Borel $\sigma$-field. Let $\theta : \mathcal{Z} \times \mathcal{Y} \to [-\infty, \infty)$ be such that*

- *$\theta$ is jointly measurable*
- *$\theta$ is not identically equal to $-\infty$ and*
- *$y \mapsto \theta(z, y)$ is upper semicontinuous for all $z \in \mathcal{Z}$.*

*In this case, one says that $-\theta$ is* normal. *Suppose in addition that there exist some $F_1 \in L$ and some $u_1 \in L^1(\mu)$ such that $\theta(z, F_1(z)) \geq u_1(z)$ for $\mu$-almost every $z$ in $\mathcal{Z}$. Then, $z \mapsto \sup_{y\in\mathcal{Y}} \theta(z, y)$ is measurable and*

$$\sup_{F\in L} \int_{\mathcal{Z}} \theta(z, F(z))\, \mu(dz) = \int_{\mathcal{Z}} \sup_{y\in\mathcal{Y}} \theta(z, y)\, \mu(dz) \in (-\infty, \infty].$$

**Definition 4.19.** The space $L$ is said to be *decomposable* if, whenever $F$ belongs to $L$ and $F_o : \mathcal{Z}_o \to \mathcal{Y}$ is a bounded measurable function on a measurable set $\mathcal{Z}_o \subset \mathcal{Z}$ of finite measure, the function $z \mapsto \mathbf{1}_{z\in\mathcal{Z}_o} F_o(z) + \mathbf{1}_{z\notin\mathcal{Z}_o} F(z)$ also belongs to $L$.

In order to obtain (4.17), we would like to apply this lemma with

- $\mathcal{Y} = C_\mathcal{X}$ equipped with the topology of uniform convergence,
- $\theta(z, f) = \langle f, q^z\rangle - \sup_{x\in\mathcal{X}}\{f(x) - J_z(x)\}$ and
- $L = C_b(\mathcal{Z}, C_\mathcal{X}) \simeq C_{\mathcal{Z}\mathcal{X}}$.

Unfortunately, two troubles occur.
*Trouble 1*: If $\mathcal{X}$ is not compact, $\mathcal{Y} = C_\mathcal{X}$ is *not* separable and fails to be a Polish space as required in the lemma. On the other hand, if $\mathcal{X}$ is compact, $C_\mathcal{X}$ is Polish.
*Trouble 2*: The space $C_{\mathcal{Z}\mathcal{X}} \simeq C_b(\mathcal{Z}, C_\mathcal{X})$ is *not* decomposable. On the other hand, the space $B(\mathcal{Z}, C_\mathcal{X})$ of all bounded and measurable functions $F : z \in \mathcal{Z} \mapsto F_z \in C_\mathcal{X}$ is decomposable.
Note that when $\mathcal{X}$ is compact, as $C_\mathcal{X}$ is separable, we have $B(\mathcal{Z}, C_\mathcal{X}) \simeq \mathcal{F}$ where $\mathcal{F}$ is the space of all the functions on $\mathcal{Z}\times\mathcal{X}$ which are bounded, $x$-continuous and $z$-measurable; such functions are jointly measurable.
We are going to apply Lemma 4.18 with

- $\mathcal{Y} = C_\mathcal{X}$ and $\mathcal{X}$ a compact Polish set,
- $\theta(z, f) = \langle f, q^z\rangle - \sup_{x\in\mathcal{X}}\{f(x) - J_z(x)\}$ for all $z \in \mathcal{Z}$ and $f \in C_\mathcal{X}$, where $q \in \mathcal{P}_{\mathcal{Z}\mathcal{X}}$ is fixed and satisfies (4.15) and
- $L = \mathcal{F}$.

As $f \mapsto \theta(z, f)$ is continuous for all $z$ and $z \mapsto \theta(z, f)$ is measurable for all $f$, $\theta$ is jointly measurable. Taking $f = 0$ gives $\theta(z, 0) = 0 > -\infty$ for all $z$, so that $\theta$ shares all the normality conditions of the lemma.
Choosing the functions $F_1 = 0 \in L$ and $u_1 = 0 \in L^1(\mu)$ leads us to $0 = \theta(z, F_1(z)) \geq u_1(z) = 0$ for every $z$ in $\mathcal{Z}$.

Therefore, we have shown that all the assumptions of Lemma 4.18 are met so that

$$\begin{aligned}&\sup_{F\in\mathcal{F}}\int_\mathcal{Z}[\langle F_z, q^z\rangle - \sup_{x\in\mathcal{X}}\{F_z(x) - J_z(x)\}]\,\mu(dz)\\ &= \int_\mathcal{Z}\sup_{f\in C_\mathcal{X}}[\langle f, q^z\rangle - \sup_{x\in\mathcal{X}}\{f(x) - J_z(x)\}]\,\mu(dz)\end{aligned} \tag{4.20}$$

whenever $\mathcal{X}$ is a compact Polish space.

To obtain (4.17), it remains to prove that for all $q$ with $q_\mathcal{Z} = \mu$,

$$\lambda^*(q) = \sup_{F\in\mathcal{F}}\int_\mathcal{Z}[\langle F_z, q^z\rangle - \sup_{x\in\mathcal{X}}\{F_z(x) - J_z(x)\}]\,\mu(dz) \tag{4.21}$$

Let us prove it without assuming that $\mathcal{X}$ is compact. Rather than invoking an abstract approximation argument, we present a specific proof of (4.21). Rewriting the above proof of Theorem 4.9 with $C_{\mathcal{Z}\mathcal{X}}$ replaced with the space $B(\mathcal{Z}\times\mathcal{X})$ of bounded measurable functions on $\mathcal{Z}\times\mathcal{X}$ one gets the following result.

**A variant of Theorem 4.9.** *Assuming (2) and (3) of Theorem 4.9, if Assumption (1) is strengtnened by "$(\mu_n)_{n\geq1}$ converges to $\mu$ in $\mathcal{P}_\mathcal{Z}$ for the stronger topology $\sigma(\mathcal{P}_\mathcal{Z}, B(\mathcal{Z}))$", then $\{K^k_n\}_{k,n\geq1}$ obeys the $(k, n)$-LDP in $\mathcal{P}_{\mathcal{Z}\mathcal{X}}$ with the topology $\sigma(\mathcal{P}_{\mathcal{Z}\mathcal{X}}, B(\mathcal{Z}\times\mathcal{X}))$ and the rate function $\tilde{i}(q) = \sup_{F\in B(\mathcal{Z}\times\mathcal{X})}\int_\mathcal{Z}[\langle F_z, q^z\rangle - \sup_{x\in\mathcal{X}}\{F_z(x) - J_z(x)\}]\,\mu(dz)$, if $q \in \mathcal{P}_{\mathcal{Z}\mathcal{X}}$ satisfies $q_\mathcal{Z} = \mu$ and $\tilde{i}(q) = +\infty$ otherwise.*

For any $\mu \in \mathcal{P}_\mathcal{Z}$, there exists a sequence of empirical measures $(\mu_n)_{n\geq1}$ as in (3.1) which converges to $\mu$ with respect to the topology $\sigma(\mathcal{P}_\mathcal{Z}, B(\mathcal{Z}))$. This can be seen as a consequence of the almost sure convergence, as $n$ tends to infinity, of the empirical measures $\frac{1}{n}\sum_{1\leq i\leq n}\delta_{Z_i}$ of the $\mu$-iid sequence of $\mathcal{Z}$-valued random variables $(Z_i)_{i\geq1}$ towards $\mu$ for the topology $\sigma(\mathcal{P}_\mathcal{Z}, B(\mathcal{Z}))$ which in turns is a corollary of the strenghened version of Sanov's theorem with the topology $\sigma(\mathcal{P}_{\mathcal{Z}}, B(\mathcal{Z}))$ on a Polish space $\mathcal{Z}$. With such a sequence $(\mu_n)_{n\geq 1}$, by Theorem 4.9 and its variant, $\{K_n^k\}_{k,n\geq 1}$ obeys the $(k,n)$-LDP in $\mathcal{P}_{\mathcal{Z}\mathcal{X}}$ with the rate functions $\lambda^*$ and $\tilde{i}$. As the rate function of a LDP is unique in a regular space (for the double index version of this known result, see [14]), we have $\lambda^* = \tilde{i}$. It follows that for all $q$ with $q_{\mathcal{Z}} = \mu$,

$$\begin{aligned}\lambda^*(q) &:= \sup_{F\in C_{\mathcal{Z}\mathcal{X}}} \int_{\mathcal{Z}} [\langle F_z, q^z\rangle - \sup_{x\in\mathcal{X}}\{F_z(x) - J_z(x)\}]\,\mu(dz)\\ &= \sup_{F\in B(\mathcal{Z}\times\mathcal{X})} \int_{\mathcal{Z}} [\langle F_z, q^z\rangle - \sup_{x\in\mathcal{X}}\{F_z(x) - J_z(x)\}]\,\mu(dz)\end{aligned}$$

which implies the desired equality (4.21).

Thanks to (4.20) and (4.21), we have proved (4.17) whenever $\mathcal{X}$ is compact. Nevertheless, the identity (4.17) will not be used directly. We shall only use (4.21) and a variant of (4.20).

Now, we have to tackle the problem of relaxing the requirement that $\mathcal{X}$ is compact. Let us take advantage of the tightness of $q$ (it is a probability on a Polish space). This means that there exists an increasing sequence $(\mathcal{K}_n^q)_{n\geq 1}$ of compact subsets of $\mathcal{Z}\times\mathcal{X}$ such that $q(\mathcal{K}_n^q) \geq 1 - 1/n$ for all $n \geq 1$. As a continuous image of a compact set, $\mathcal{X}_n^q := \{x \in \mathcal{X}; (z, x) \in \mathcal{K}_n^q$ for some $z \in \mathcal{Z}\}$ is a compact set. We also have $q(\mathcal{Z} \times \mathcal{X}_n^q) \geq 1 - 1/n$ for all $n$. It follows that for $q_{\mathcal{Z}}$-almost every $z \in \mathcal{Z}$, $q^z$ is determined by the values $\langle f, q^z\rangle$ where $f$ describes the set $\bigcup_{n\geq 1} \mathcal{C}(\mathcal{X}_n^q)$ where, for any measurable set $\mathcal{X}_o$ in $\mathcal{X}$, we denote

$$\mathcal{C}(\mathcal{X}_o) = \mathbf{1}_{\mathcal{X}_o} C_{\mathcal{X}} = \{f : \mathcal{X} \to \mathbb{R}; f = \mathbf{1}_{\mathcal{X}_o}\tilde{f}, \text{ for some } \tilde{f} \in C_{\mathcal{X}}\}. \tag{4.22}$$

To see this, remark that for all measurable set $A$ in $\mathcal{X}$ such that $A \cap (\cup_n \mathcal{X}_n^q) = \emptyset$, we have $\int_{\mathcal{Z}} q^z(A)\,\mu(dz) = q(\mathcal{Z} \times A) = \lim_{n\to\infty} q(\mathcal{Z} \times (A \cap \mathcal{X}_n^q)) = 0$.

We can now proceed with the proof of $\lambda^*(q) \geq i(q)$ for all $q$ satisfying (4.15). For all such $q$ we have,

$$\begin{aligned}\lambda^*(q) &= \sup_{F\in\mathcal{F}} \int_{\mathcal{Z}} [\langle F_z, q^z\rangle - \sup_{x\in\mathcal{X}}\{F_z(x) - J_z(x)\}]\,\mu(dz)\\ &\overset{(a)}{\geq} \sup_{n\geq 1}\ \sup_{F\in B\left(\mathcal{Z},\mathcal{C}(\mathcal{X}_n^q)\right)} \int_{\mathcal{Z}} [\langle F_z, q^z\rangle - \sup_{x\in\mathcal{X}}\{F_z(x) - J_z(x)\}]\,\mu(dz)\\ &\overset{(b)}{=} \sup_{n\geq 1} \int_{\mathcal{Z}} \sup_{f\in\mathcal{C}(\mathcal{X}_n^q)} [\langle f, q^z\rangle - \sup_{x\in\mathcal{X}}\{f(x) - J_z(x)\}]\,\mu(dz)\\ &\overset{(c)}{=} \int_{\mathcal{Z}} \sup_{f\in\bigcup_{n\geq 1}\mathcal{C}(\mathcal{X}_n^q)} [\langle f, q^z\rangle - \sup_{x\in\mathcal{X}}\{f(x) - J_z(x)\}]\,\mu(dz)\\ &\overset{(d)}{=} \int_{\mathcal{Z}\times\mathcal{X}} J_z(x)\, q^z(dx)\mu(dx)\\ &= i(q)\end{aligned}$$

where the first equality is (4.21). The remaining series of inequality and equalities needs to be justified. This will require two more lemmas the proofs of which are postponed after the proof of the present proposition.

• Inequality (a). It is enough to show that for any function $F \in B(\mathcal{Z}, \mathcal{C}(\mathcal{X}_o))$ with $\mathcal{X}_o$ a compact subset of $\mathcal{X}$, there exists a sequence $(F^n)_{n\geq 1}$ in $\mathcal{F}$ such that

$$\begin{aligned}(4.23)\qquad & \int_{\mathcal{Z}} [\langle F_z, q^z\rangle - \sup_{x\in\mathcal{X}}\{F_z(x) - J_z(x)\}]\,\mu(dz)\\ &= \lim_{n\to\infty}\int_{\mathcal{Z}}[\langle F^n_z, q^z\rangle - \sup_{x\in\mathcal{X}}\{F^n_z(x) - J_z(x)\}]\,\mu(dz).\end{aligned}$$

Let us show that

$$F^n_z(x) := \sup\{F_z(y) - nd(x,y); y\in\mathcal{X}\}$$

does this job. For each $z$, $(-F^n_z)_{n\geq 1}$ is the Moreau-Yosida approximation of $-F_z$, and it is a well-known result (see [4], Section 1.7.3 for instance) that

- for all $z\in\mathcal{Z}$, $x\mapsto F^n_z(x)$ is $n$-Lipschitz,

and for all $(z,x)\in\mathcal{Z}\times\mathcal{X}$,

- $-\|F\| \leq F_z(x) \leq F^n_z(x) \leq \|F_z\| \leq \|F\|$, where $\|\cdot\|$ stands for the uniform norm,
- $(F^n_z(x))_{n\geq 1}$ is a decreasing sequence and
- $\lim_{n\to\infty} F^n_z(x) = F_z(x)$

For the last statement, note that it is necessary that $F_z$ is upper semicontinuous on $\mathcal{X}$. But, this is insured by the assumption that $\mathcal{X}_o$ is closed and $F_z\in\mathcal{C}(\mathcal{X}_o)$.

Now let us make sure that for any $x_o\in\mathcal{X}$, $z\mapsto F^n_z(x_o)$ is measurable. For all real $a$, we have

$$\begin{aligned} F^n_z(x_o)\leq a &\Leftrightarrow \forall y\in\mathcal{X}, F_z(y) - nd(x_o,y)\leq a\\ &\Leftrightarrow \forall k\geq 1, F_z(x_k) - nd(x_o,x_k)\leq a\end{aligned}$$

where $\{x_k; k\geq 1\}$ is a countable dense subset of $\mathcal{X}$ (recall that $\mathcal{X}$ is Polish). This holds, since $y\mapsto F_z(y) - nd(x_o,y)$ is continuous. It follows that $\{z\in\mathcal{Z}; F^n_z(x_o)\leq a\} = \cap_{k\geq 1}\{z\in\mathcal{Z}; F_z(x_k) - nd(x_o,x_k)\leq a\}$. As $z\mapsto F_z(x_k)$ is measurable for all $k$, this proves the measurability of $z\mapsto F^n_z(x_o)$. Therefore, $F^n$ belongs to $\mathcal{F}$ for all $n\geq 1$.

With the estimate $-\|F\|\leq F_z(x)\leq F^n_z(x)\leq \|F_z\|\leq\|F\|$ and the limit $\lim_{n\to\infty}F^n_z(x) = F_z(x)$, one can apply the dominated convergence theorem to obtain that

$$(4.24)\qquad \lim_{n\to\infty}\int_{\mathcal{Z}}\langle F^n_z, q^z\rangle\,\mu(dz) = \lim_{n\to\infty}\int_{\mathcal{Z}\times\mathcal{X}} F^n\,dq = \int_{\mathcal{Z}\times\mathcal{X}} F\,dq = \int_{\mathcal{Z}}\langle F_z, q^z\rangle\,\mu(dz).$$

Similarly, the limit

$$(4.25)\qquad \lim_{n\to\infty}\int_{\mathcal{Z}}\sup_{x\in\mathcal{X}}\{F^n_z(x) - J_z(x)\}\,\mu(dz) = \int_{\mathcal{Z}}\sup_{x\in\mathcal{X}}\{F_z(x) - J_z(x)\}\,\mu(dz)$$

follows from the estimate (4.8) and the following lemma.

**Lemma 4.26.** *Let $J$ be an inf-compact $[0,\infty]$-valued function on $\mathcal{X}$ and $(f_n)_{n\geq 1}$ a decreasing sequence of continuous bounded functions on $\mathcal{X}$ which converges pointwise to some bounded upper semicontinuous function $f$. Then, $(\sup_{x\in\mathcal{X}}\{f_n(x) - J(x)\})_{n\geq 1}$ is a decreasing sequence and*

$$\lim_{n\to\infty}\sup_{x\in\mathcal{X}}\{f_n(x) - J(x)\} = \sup_{x\in\mathcal{X}}\{f(x) - J(x)\}.$$

The proof of this lemma is put back after the proof of the present proposition.

Finally, (4.23) follows from (4.24) and (4.25).

• Equality (b) is a variant of (4.20) applied with the compact set $\mathcal{X}^q_n$.

• Equality (c). If the sequence $\mathcal{C}(\mathcal{X}_n^q)_{n\geq 1}$ were increasing, equality (c) would be a direct consequence of the monotone convergence theorem. Nevertheless, this is almost the case since, for any pair of closed subsets $\mathcal{X}_o$ and $\mathcal{X}_1$ of $\mathcal{X}$ such that $\mathcal{X}_o \subset \mathcal{X}_1$, any function $f \in \mathcal{C}(\mathcal{X}_o)$ can be approximated pointwise by a uniformly bounded decreasing sequence $(f_n)$ in $\mathcal{C}(\mathcal{X}_1)$ such that $\lim_{n\to\infty} \sup_{x\in\mathcal{X}}\{f_n(x) - J_z(x)\} = \sup_{x\in\mathcal{X}}\{f(x) - J_z(x)\}$. One proves this, exactly as for inequality (a), by means of a Moreau-Yosida approximation and Lemma 4.26. With this in hand, equality (c) follows from the monotone convergence theorem.

• Equality (d). This equality is a consequence of the following lemma.

**Lemma 4.27.** *Let $J$ be a $[0,+\infty]$-valued lower semicontinuous function on $\mathcal{X}$.*

*If $C_{\mathcal{X}}$ in Lemma 4.16 is replaced with the set $\mathcal{G}_Q = \bigcup_{n\geq 1} \mathcal{C}(\mathcal{X}_n^Q)$ where $(\mathcal{X}_n^Q)_{n\geq 1}$ is an increasing sequence of closed subsets of $\mathcal{X}$ such that $\lim_{n\to\infty} Q(\mathcal{X}_n^Q) = 1$, then we still have*

$$\sup_{f\in\mathcal{G}_Q} \left\{ \int_{\mathcal{X}} f\, dQ - \sup_{x\in\mathcal{X}}(f(x) - J(x)) \right\} = \int_{\mathcal{X}} J\, dQ.$$

The proof of this lemma is put back after the proof of the present proposition.

Note that we have already remarked that for $q_{\mathcal{Z}}$-almost every $z \in \mathcal{Z}$, $q^z$ is determined by the values $\langle f, q^z\rangle$ where $f$ describes the set $\bigcup_{n\geq 1} \mathcal{C}(\mathcal{X}_n^q)$. One obtains equality (d) by means of Lemma 4.27, with $\mathcal{G}_{q^z} = \bigcup_{n\geq 1} \mathcal{C}(\mathcal{X}_n^q)$, for all $z \in \mathcal{Z}$.

We have proved that $\lambda^* \geq i$ and this completes the proof of the proposition. □

*A comment on this proof.* One could think of replacing the spaces $\mathcal{C}(\mathcal{X}_n^q)$ defined by (4.22) with the smaller spaces $\hat{\mathcal{C}}(\mathcal{X}_n^q)$ of continuous functions on $\mathcal{X}$ with their support in $\mathcal{X}_n^q$. This clearly provides an increasing sequence and simplifies the proof of equality (c). But unfortunately, equality (a) doesn't work anymore since $\hat{\mathcal{C}}(\mathcal{X}_o)$ reduces to the null space when the compact set $\mathcal{X}_o$ has an empty interior (a common feature in infinite dimension).

## 4.4. **Proofs of the lemmas.** 

We go on with the proofs of Lemmas 4.16 , 4.26 and 4.27.

*Proof of Lemmas 4.16 and 4.27.* Lemma 4.16 is a particular case of Lemma 4.27, we only prove Lemma 4.27.

As, $\sup_{R\in\mathcal{P}_{\mathcal{X}}} \langle f-J, R\rangle \leq \sup_{x\in\mathcal{X}}\{f(x)-J(x)\} = \sup_{x\in\mathcal{X}}\langle f-J, \delta_x\rangle \leq \sup_{R\in\mathcal{P}_{\mathcal{X}}}\langle f-J, R\rangle$, we have $\sup_{x\in\mathcal{X}}\{f(x) - J(x)\} = \sup_{R\in\mathcal{P}_{\mathcal{X}}}\langle f - J, R\rangle$. Therefore,

$$\begin{aligned}
\sup_{f\in\mathcal{G}_Q} \left\{ \langle f, Q\rangle - \sup_{x\in\mathcal{X}}\{f(x) - J(x)\} \right\} &= \sup_{f\in\mathcal{G}_Q} \left\{ \langle f, Q\rangle - \sup_{R\in\mathcal{P}_{\mathcal{X}}} \langle f - J, R\rangle \right\} \\
&= \langle J, Q\rangle + \sup_{f\in\mathcal{G}_Q} \left\{ \langle f - J, Q\rangle - \sup_{R\in\mathcal{P}_{\mathcal{X}}} \langle f - J, R\rangle \right\} \\
&\leq \langle J, Q\rangle
\end{aligned}$$

where the last inequality holds since $Q \in \mathcal{P}_{\mathcal{X}}$.

Now, let's prove the converse inequality. As $J$ is a lower semicontinuous function which is bounded below, it is the pointwise limit of an increasing sequence $(\tilde{J}_n)_{n\geq 1}$ in $C_{\mathcal{X}}$: once again, the Moreau-Yosida approximation: $\tilde{J}_n(x) = \inf\{J(y) + nd(x, y); y \in \mathcal{X}\}$.

Let us define $J_n(x) = \mathbf{1}_{\mathcal{X}_n^Q}(x)(0 \vee \tilde{J}_n(x) \wedge n)$ for all $x$ and $n$. As $(\mathcal{X}_n^Q)_{n\geq 1}$ is an increasing sequence of sets, $(J_n)_{n\geq 1}$ is an increasing sequence of functions such that for all $n$, $J_n$ is in $\mathcal{C}(\mathcal{X}_n^Q)$. We have

$$\begin{aligned}
\sup_{f\in\mathcal{G}_Q}\left\{\langle f, Q\rangle - \sup_{x\in\mathcal{X}}\{f(x) - J(x)\}\right\} &\overset{(a)}{\geq} \sup_{n\geq 1}\left(\langle J_n, Q\rangle - \sup_{x\in\mathcal{X}}\{J_n(x) - J(x)\}\right) \\
&\overset{(b)}{\geq} \sup_{n\geq 1}\int_{\mathcal{X}} J_n\, dQ \\
&\overset{(c)}{\geq} \sup_{k\geq 1}\sup_{n\geq k}\int_{\mathcal{X}_k^Q}(0\vee \tilde{J}_n \wedge n)\, dQ, \\
&\overset{(d)}{=} \sup_{k\geq 1}\int_{\mathcal{X}_k^Q} J\, dQ, \\
&\overset{(e)}{=} \int_{\mathcal{X}} J\, dQ
\end{aligned}$$

where inequality (a) holds since $J_n \in \mathcal{C}(\mathcal{X}_n^Q)$, inequality (b) follows from $J_n \leq J$, equality (c) holds since the sequence $(\mathcal{X}_n^Q)$ is increasing, equality (d) follows from the monotone convergence theorem and equality (e) follows from the monotone convergence theorem together with $\lim_{k\to\infty} Q(\mathcal{X}\setminus\mathcal{X}_k^Q) = 0$. This completes the proof of the lemmas. ☐

*Proof of Lemma 4.26.* Changing sign and denoting $g_n(x) = J(x) - f_n(x)$, $g(x) = J(x) - f(x)$, we want to prove that $\lim_{n\to\infty}\inf_{x\in\mathcal{X}} g_n(x) = \inf_{x\in\mathcal{X}} g(x)$.

We see that $(g_n)_{n\geq 1}$ is an increasing sequence of lower semicontinuous functions. It follows by the Proposition 5.4 of [15] that it is a $\Gamma$-convergent sequence and

$$\Gamma\text{-}\lim_{n\to\infty} g_n = \lim_{n\to\infty} g_n = g. \tag{4.28}$$

Let us admit for a while that there exists some compact set $K$ which satisfies

$$\inf_{x\in\mathcal{X}} g_n(x) = \inf_{x\in K} g_n(x) \tag{4.29}$$

for all $n$. This and the convergence (4.28) allows to apply Theorem 7.4 of [15] to obtain $\lim_{n\to\infty}\inf_{x\in\mathcal{X}} g_n(x) = \inf_{x\in\mathcal{X}}\Gamma\text{-}\lim_{n\to\infty} g_n(x) = \inf_{x\in\mathcal{X}} g(x)$ which is the desired result.

It remains to check that (4.29) is true. Let $x_* \in \mathcal{X}$ be such that $J(x_*) < \infty$ (if $J \equiv +\infty$, there is nothing to prove). Then, $\inf_{x\in\mathcal{X}} g_n(x) \leq g_n(x_*) = J(x_*) - f_n(x_*) \leq J(x_*) - f(x_*) \leq J(x_*) - \inf_{x\in\mathcal{X}} f(x)$. On the other hand, for all $x$ and $n$, $f_n(x) \leq f_1(x) \leq A := \sup f_1$. Let $B := A + 1 + J(x_*) - \inf_{x\in\mathcal{X}} f(x)$. For all $x$ such that $J(x) > B$, we have $g_n(x) > B - \sup_{x\in\mathcal{X}} f_n(x) \geq B - A \geq J(x_*) - \inf_{x\in\mathcal{X}} f(x) + 1$. We have just seen that for all $n$,

$$\begin{aligned}
\inf_{x\in\mathcal{X}} g_n(x) &\leq J(x_*) - \inf_{x\in\mathcal{X}} f(x) \\
\inf_{x; J(x)>B} g_n(x) &\geq J(x_*) - \inf_{x\in\mathcal{X}} f(x) + 1
\end{aligned}$$

This proves (4.29) with the compact level set $K = \{J \leq B\}$ and completes the proof of the lemma. ☐

## 5. Large deviations of a doubly indexed sequence of random measures. Main results

Theorem 4.9 states a $(k,n)$-LDP for $K_n^k = \frac{1}{n}\sum_{i=1}^n \delta_{(z_{n,i},X_{n,i}^k)}$ but we are mostly interested in the $(k,n)$-LD in $\mathcal{P}_\mathcal{X}$ of $L_n^k = \frac{1}{n}\sum_{i=1}^n \delta_{X_{n,i}^k}$. It will easily follow from Theorem 4.9 and the contraction principle. Let us denote

$$P^k(dx) = \int_\mathcal{Z} P_z^k(dx)\,\mu(dz) \in \mathcal{P}_\mathcal{X}, k \geq 1.$$

**Theorem 5.1.** *Suppose that*

(1) $(\mu_n)_{n\geq 1}$ *converges to* $\mu$ *in* $\mathcal{P}_\mathcal{Z}$,
(2) *for each* $k \geq 1$, $(P_z^k; z \in \mathcal{Z})$ *is a Feller system in the sense of Definition 2.8,*
(3) *for each* $z \in \mathcal{Z}$, $(P_z^k)_{k\geq 1}$ *obeys the k-LDP in* $\mathcal{X}$ *with the good rate function* $J_z$.

*Then the following statements hold true.*

(a) $\{L_n^k\}_{k,n\geq 1}$ *obeys the* $(k,n)$*-LDP in* $\mathcal{P}_\mathcal{X}$ *with the good rate function* $I$ *which is defined for all* $Q \in \mathcal{P}_\mathcal{X}$ *by*

$$I(Q) = \inf\left\{\int_{\mathcal{Z}\times\mathcal{X}} J_z(x)\,\mu(dz)\Pi_z(dx); (\Pi_z)_{z\in\mathcal{Z}} : \int_\mathcal{Z} \Pi_z\,\mu(dz) = Q\right\} \tag{5.2}$$

*where the transition kernels* $z \in \mathcal{Z} \mapsto \Pi_z \in \mathcal{P}_\mathcal{X}$ *are measurable.*

(b) *Another representation of this rate function is*

$$I(Q) = \sup_{f\in C_\mathcal{X}} \left\{\int_\mathcal{Z} \mathcal{S}f(z)\,\mu(dz) - \int_\mathcal{X} f(x)\,Q(dx)\right\}, Q \in \mathcal{P}_\mathcal{X}$$

*where* $\mathcal{S}f(z)$ *is defined for all* $z \in \mathcal{Z}$ *by*

$$\mathcal{S}f(z) = \inf_{x\in\mathcal{X}}\{J_z(x) + f(x)\}.$$

(c) *If* $I(Q) < +\infty$, *there exists a (possibly not unique) kernel* $(\Pi_z^*)_{z\in\mathcal{Z}}$ *which realizes the infimum in (5.2).*

(d) *If for each* $k$ *the Feller system* $(P_z^k)_{z\in\mathcal{Z}}$ *satisfies*

$$P_z^k = P^k(\cdot \mid \beta(X) = z) \tag{5.3}$$

*for* $\mu$*-almost every* $z \in \mathcal{Z}$ *and some continuous function* $\beta : \mathcal{X} \to \mathcal{Z}$, *we have*

$$I(Q) = \begin{cases} \int_\mathcal{X} J_{\beta(x)}(x)\,Q(dx) & \textit{if } \beta \diamond Q = \mu \\ +\infty & \textit{otherwise} \end{cases}, \quad Q \in \mathcal{P}_\mathcal{X}. \tag{5.4}$$

The dual space $C'_\mathcal{X}$ of $(C_\mathcal{X}, \|\cdot\|)$ is equipped with the $*$-weak topology $\sigma(C'_\mathcal{X}, C_\mathcal{X})$, see Section 1.7.

*Proof.* Let us prove (a). As $L_n^k$ is the $\mathcal{X}$-marginal of $K_n^k$ and $\{K_n^k\}$ obeys the $(k,n)$-LDP with the good rate function $i$, the statement (a) follows from an obvious extension to the double index setting of the contraction principle (see [14]): $\{L_n^k\}$ obeys the $(k,n)$-LDP in $\mathcal{P}_\mathcal{X}$ with the good rate function

$$I(Q) = \inf\{i(q); q \in \mathcal{P}_{\mathcal{Z}\mathcal{X}} : q_\mathcal{X} = Q\} \tag{5.5}$$

which is (5.2).

Let us prove (b). We rewrite the proof of Theorem 4.9 with $L_n^k$ instead of $K_n^k$. As in the proof of Proposition 3.14, we replace $F(z,x)$ by $f(x)$ to obtain the pointwise convergence of the normalized log-Laplace transforms

$$\lim_{k\to\infty} \Lambda_k(f) = \Lambda(f) \tag{5.6}$$

for all $f \in C_{\mathcal{X}}$, with

$$\begin{aligned}\Lambda_k(f) &= \frac{1}{k}\int_{\mathcal{Z}} \log\langle e^{kf}, P_z^k\rangle\,\mu(dz) \quad \text{and}\\ \Lambda(f) &= \int_{\mathcal{Z}} \sup_{x\in\mathcal{X}}\{f(x) - J_z(x)\}\,\mu(dz).\end{aligned}$$

Note that $\Lambda_k(f) = \lambda_k(F_f)$ and $\Lambda(f) = \lambda(F_f)$ with $F_f(z,x) = f(x)$, so that (5.6) is a specialization of (4.6).
Exactly the same arguments as in the proof of Theorem 4.9 allow us to establish that $\{L_n^k\}_{k,n\geq 1}$ obeys the LDP in $C'_{\mathcal{X}}$ with the rate function $\Lambda^*(Q) = \sup_{f\in C_{\mathcal{X}}}\{\langle f, Q\rangle - \Lambda(f)\}$, $Q \in C'_{\mathcal{X}}$. In particular, (4.7) and (4.11) become

$$|\Lambda_k(f)| \leq \|f\|, \quad |\Lambda(f)| \leq \|f\|, \tag{5.7}$$

for all $f \in C_{\mathcal{X}}$ and

$$\Gamma\text{-}\lim_{k\to\infty} \Lambda_k^* = \Lambda^* \tag{5.8}$$

in $C'_{\mathcal{X}}$, where these convex conjugates are taken with respect to the duality $(C'_{\mathcal{X}}, C_{\mathcal{X}})$.

Thanks to (4.12), (5.5) and the uniqueness of the rate function (see [14]), we see that $\{\Lambda^* < +\infty\} \subset \mathcal{P}_{\mathcal{X}}$. We conclude as in the proof of Theorem 4.9 that $\{L_n^k\}_{k,n\geq 1}$ obeys the LDP in $\mathcal{P}_{\mathcal{X}}$ with the rate function $\Lambda^*(Q) = \sup_{f\in C_{\mathcal{X}}}\{\langle f, Q\rangle - \Lambda(f)\}$, $Q \in \mathcal{P}_{\mathcal{X}}$. As the rate function is unique,

$$I = \Lambda^*. \tag{5.9}$$

Considering $-f$ instead of $f$ in $\sup_{f\in C_{\mathcal{X}}}$ leads to statement (b).

Let us prove (c). As $i$ is a good rate function, the result follows from the identity (5.5).

Finally, statement (d) is a direct consequence of Lemma 5.13 below. $\square$

Let us introduce the $[0,\infty]$-valued functions $I$ and $I_k$ on $C'_{\mathcal{X}}$ which are defined for all $k \geq 1$ and $Q \in C'_{\mathcal{X}}$ by

$$I_k(Q) = \inf\left\{\int_{\mathcal{Z}} \frac{1}{k} H(q^z | P_z^k)\,\mu(dz); q \in \mathcal{P}_{\mathcal{Z}\mathcal{X}} : q_{\mathcal{Z}} = \mu, q_{\mathcal{X}} = Q\right\} \tag{5.10}$$

$$I(Q) = \inf\left\{\int_{\mathcal{Z}\times\mathcal{X}} J_z(x)\,q(dzdx); q \in \mathcal{P}_{\mathcal{Z}\mathcal{X}} : q_{\mathcal{Z}} = \mu, q_{\mathcal{X}} = Q\right\} \tag{5.11}$$

where we use the same notation $I(Q)$ for the function on $U_{\mathcal{X}}$ and its restriction to $\mathcal{P}_{\mathcal{X}}$ (see (5.5)) and the convention that $\inf\emptyset = +\infty$. In particular, the effective domains of $I_k$ and $I$ are included in $\mathcal{P}_{\mathcal{X}}$.

As a by-product of the proof of Theorem 5.1, we have the following corollary.

**Corollary 5.12.** *[Hypotheses of Theorem 5.1]. The sequence $(I_k)_{k\geq 1}$ $\Gamma$-converges to $I$ in $C'_{\mathcal{X}}$.*

*Proof.* We have shown at (5.9) that $\Lambda^* = I$. It is also true that $\Lambda_k^* = I_k$, as can be shown by a minor modification of the proof of (3.17). One concludes with (5.8). $\square$

During the proof of Theorem 5.1, we have invoked the following

**Lemma 5.13.** *[Hypotheses of Theorem 5.1]. If for each $k$ the Feller system $(P_z^k)_{z\in\mathcal{Z}}$ satisfies (5.3) with $\beta$ continuous, $I$ is given by*

$$I(Q) = \begin{cases} \int_{\mathcal{X}} J_{\beta(x)}(x)\,Q(dx) & \textit{if } Q\in\mathcal{P}_{\mathcal{X}} \textit{ and } \beta\diamond Q=\mu \\ +\infty & \textit{otherwise} \end{cases}, \quad Q\in U_{\mathcal{X}}. \tag{5.14}$$

*Proof.* Let us first show that $\operatorname{dom} I$ is included in $P_\beta(\mu) := \{Q\in C'_{\mathcal{X}}; Q\in\mathcal{P}_{\mathcal{X}}, \beta\diamond Q=\mu\}$, whenever $\beta$ is continuous.
As a direct consequence of Proposition 3.14-c, we obtain for all $Q\in C'_{\mathcal{X}}$ that

$$I_k(Q) = \begin{cases} \frac{1}{k}H(Q|P^k) & \text{if } Q\in\mathcal{P}_{\mathcal{X}} \text{ and } \beta\diamond Q=\mu \\ +\infty & \text{otherwise} \end{cases}. \tag{5.15}$$

This holds with $\beta$ measurable, see Remark 3.19. Hence, $\operatorname{dom} I_k \subset P_\beta(\mu)$ for each $k$. Corollary 5.12 implies that $\operatorname{dom} I$ is included in the closure of $P_\beta(\mu)$ in $C'_{\mathcal{X}}$. As $\beta$ is assumed to be continuous, $\{Q\in C'_{\mathcal{X}}; \langle Q, g\circ\beta\rangle = \langle\mu,g\rangle, \forall g\in C_{\mathcal{Z}}\}$ is closed in $C'_{\mathcal{X}}$ and one obtains the inclusion $\operatorname{dom} I\subset\{Q\in C'_{\mathcal{X}}; \langle Q, g\circ\beta\rangle = \langle\mu,g\rangle, \forall g\in C_{\mathcal{Z}}\}$. On the other hand, $\operatorname{dom} I\subset\mathcal{P}_{\mathcal{X}}$. Therefore, we obtain the desired inclusion $\operatorname{dom} I\subset\{Q\in C'_{\mathcal{X}}; \langle Q, g\circ\beta\rangle = \langle\mu,g\rangle, \forall g\in C_{\mathcal{Z}}\}\cap\mathcal{P}_{\mathcal{X}} = P_\beta(\mu)$.
This implies that (5.11) admits the unique minimizer $q^*(dzdx) = \mu(dz)Q(dx\mid\beta(X)=z)$ and gives (5.14). □

## 6. Applications to the optimal transport

We apply the main results of Sections 4 and 5 to the setting of Section 2. The space $\mathcal{X}=\mathbb{R}^{2d}$ is the space of the random couples and $\mathcal{Z}=\mathbb{R}^d$ is the space of the initial positions. The empirical random measures $N_n^k$ and $M_n^k$ are specified by (2.2), (2.4) and (2.16). In the whole present section, the Assumptions 2.7 are supposed to hold.

The spaces $C_{\mathbb{R}^d}$ and $C_{\mathbb{R}^{2d}}$ of all continuous bounded functions on $\mathbb{R}^d$ and $\mathbb{R}^{2d}$ are equipped with their topologies of uniform convergence and their dual spaces $C'_{\mathbb{R}^d}$ and $C'_{\mathbb{R}^{2d}}$ are equipped with the corresponding $*$-weak topologies, see Section 1.7. It is convenient to use the notation

$$\xi_A(y) = \begin{cases} 0 & \text{if } y\in A \\ +\infty & \text{if } y\notin A \end{cases}$$

which is called the "convex" indicator of the subset $A$ ($\xi_A$ is a convex function if and only if $A$ is a convex set). Under the Assumptions 2.7, the assumptions of Theorem 5.1 are satisfied with

$$J_z(x) = c(x_0,x_1) + \xi_{x_0=z}, \quad x=(x_0,x_1)\in\mathbb{R}^{2d}, z\in\mathbb{R}^d$$

where $c$ is given at (2.10). Let $\pi^k$ be defined by (2.17). In the present setting, the functions $I_k$ and $I$ defined at (5.10) and (5.11) are given for all $\rho\in C'_{\mathbb{R}^{2d}}$ by $I_k=S_k$ and $I=S$ where

$$\begin{aligned} S_k(\rho) &= \frac{1}{k}H(\rho|\pi^k) + \xi_{\Pi_0(\mu)}(\rho) \\ S(\rho) &= \int_{\mathbb{R}^{2d}} c\,d\rho + \xi_{\Pi_0(\mu)}(\rho) \end{aligned}$$

with $\int_{\mathbb{R}^{2d}} c\,d\rho = \int_{\mathbb{R}^{2d}} c(x_0,x_1)\,\rho(dx_0dx_1)$ and

$$\Pi_0(\mu) = \{\rho\in C'_{\mathbb{R}^{2d}}; \langle\rho,\varphi\circ X_0\rangle = \langle\mu,\varphi\rangle, \forall\varphi\in C_{\mathbb{R}^d}\}$$

and the convention that $H(\rho|\pi^k)=+\infty$ and $\int_{\mathbb{R}^{2d}} c\,d\rho=+\infty$ for all $\rho\in C'_{\mathbb{R}^{2d}}\setminus\mathcal{P}_{\mathbb{R}^{2d}}$. Of course, $\Pi_0(\mu)\cap\mathcal{P}_{\mathbb{R}^{2d}}$ is the set of all probability measures on $\mathbb{R}^{2d}$ such that $\rho_0=\mu$.
The reason for introducing $C'$ besides $\mathcal{P}$, is that the strong unit ball $U$ of $C'$ is $*$-weak

compact, while compactness in $\mathcal{P}$ requires tightness criteria. This will considerably simplify the compactness arguments.
To see that the identity about $S$ holds true, observe that the canonical projection $X_0$ is continuous. In particular, we have (5.3) with the continuous function $\beta = X_0$, which by Lemma 5.13 gives (5.14). The identity about $S_k$ is (5.15) with $\beta = X_0$.
We shall also use the sets

$$\begin{aligned}\Pi_1(\nu) &= \{\rho \in C'_{\mathbb{R}^{2d}}; \langle \rho, \varphi \circ X_1\rangle = \langle \nu, \varphi\rangle, \forall \varphi \in C_{\mathbb{R}^d}\} \quad \text{and} \\ \Pi(\mu,\nu) &= \Pi_0(\mu) \cap \Pi_1(\nu).\end{aligned}$$

As $X_0$ and $X_1$ are continuous, $\Pi_0(\mu)$ and $\Pi_1(\nu)$ are well-defined subsets of $C'_{\mathbb{R}^{2d}}$ since $\varphi \circ X_0$ and $\varphi \circ X_1$ are in $C_{\mathbb{R}^{2d}}$. We use the same notation for $\Pi(\mu,\nu)$ in $\mathcal{P}_{\mathbb{R}^{2d}}$ and $C'_{\mathbb{R}^{2d}}$.
We define for all $\nu \in \mathcal{P}_{\mathbb{R}^d}$ and all $k$

$$\begin{aligned}T_k(\nu) &= \inf_{\rho \in \Pi(\mu,\nu)} \frac{1}{k} H(\rho|\pi^k) \\ T(\nu) &= \inf_{\rho \in \Pi(\mu,\nu)} \int_{\mathbb{R}^{2d}} c\, d\rho\end{aligned}$$

and we set $T_k(\nu) = T(\nu) = +\infty$ whenever $\nu \in C'_{\mathbb{R}^d} \setminus \mathcal{P}_{\mathbb{R}^d}$.
*Caution.* We'll denote similarly the rate functions $S_k, S, T_k$ and $T$ on $C'$ and their restrictions to $\mathcal{P}$.

**Lemma 6.1.** *For each $k$,*

(a) *$\{M_n^k\}_{n\geq 1}$ obeys the $n$-LDP in $\mathcal{P}_{\mathbb{R}^{2d}}$ and $C'_{\mathbb{R}^{2d}}$ with the good rate function $kS_k$ and*
(b) *$\{N_n^k\}_{n\geq 1}$ obeys the $n$-LDP in $\mathcal{P}_{\mathbb{R}^d}$ and $C'_{\mathbb{R}^d}$ with the good rate function $kT_k$.*

*Proof.* To get (a), apply Proposition 3.14; (b) follows by the contraction principle. □

Applying Theorem 5.1, one obtains

**Theorem 6.2.** *The following assertions hold true.*

(a) *$\{N_n^k\}_{k,n\geq}$ obeys the $(k,n)$-LDP in $\mathcal{P}_{\mathbb{R}^d}$ with the rate function $\nu \in \mathcal{P}_{\mathbb{R}^d} \mapsto \mathcal{T}_c(\mu,\nu) \in [0,\infty]$.*
(b) *For all $\nu \in \mathcal{P}_{\mathbb{R}^d}$,*

$$\mathcal{T}_c(\mu,\nu) = \sup_{f \in C_{\mathbb{R}^d}} \left\{ \int_{\mathbb{R}^d} \mathcal{S}_1 f(x_0)\, \mu(dx_0) - \int_{\mathbb{R}^d} f(x_1)\, \nu(dx_1) \right\}$$

*with $\mathcal{S}_1 f(z) = \inf_{x_1 \in \mathbb{R}^d}\{c(z,x_1) + f(x_1)\}$, $z \in \mathbb{R}^d$.*

*Remark* 6.3. The statement (b) of this theorem is the *Kantorovich duality* ([24], Theorem 1.3) and Theorem 5.1-(b) is a general version of this duality result.

Similarly, we have the

**Proposition 6.4.** *The following assertions hold true.*

(a) *$\{M_n^k\}_{k,n\geq}$ obeys the $(k,n)$-LDP in $\mathcal{P}_{\mathbb{R}^{2d}}$ with the rate function $S$.*
(b) *For all $\rho \in \mathcal{P}_{\mathbb{R}^{2d}}$ such that $\rho_0 = \mu$*

$$\int_{\mathbb{R}^{2d}} c\, d\rho = \sup_{g \in C_{\mathbb{R}^{2d}}} \left\{ \int_{\mathbb{R}^d} \mathcal{S}_{01} g(x_0)\, \mu(dx_0) - \int_{\mathbb{R}^{2d}} g(x_0,x_1)\, \rho(dx_0 dx_1) \right\}$$

*with $\mathcal{S}_{01} g(z) = \inf_{x_1 \in \mathbb{R}^d}\{c(z,x_1) + g(z,x_1)\}$, $z \in \mathbb{R}^d$.*

As a consequence of the preceding results, we have the

**Theorem 6.5.** *The following assertions hold true*

(a) $\Gamma\text{-}\lim_{k\to\infty} S_k = S$ *in* $C'_{\mathbb{R}^{2d}}$ *and* $\mathcal{P}_{\mathbb{R}^{2d}}$

(b) $\Gamma\text{-}\lim_{k\to\infty} T_k = T$ *in* $C'_{\mathbb{R}^{d}}$ *and* $\mathcal{P}_{\mathbb{R}^{d}}$.

(c) *Since* $\Gamma\text{-}\lim_{k\to\infty} T_k = T$, *for all* $\nu \in \mathcal{P}_{\mathbb{R}^d}$ *there exists a sequence* $(\nu_k)$ *in* $\mathcal{P}_{\mathbb{R}^d}$ *such that* $\lim_{k\to\infty} \nu_k = \nu$ *in* $\mathcal{P}_{\mathbb{R}^d}$ *and* $\lim_{k\to\infty} T_k(\nu_k) = T(\nu)$ *in* $[0,\infty]$.

*Proof.* It is proved in [14] that in a Polish space $\mathcal{X}$, if one has a $k$-indexed family of $n$-LDPs with rate functions $kI_k$ such that the doubly indexed sequence obeys the (weak) $(k,n)$-LDP with rate function $I$, then $\Gamma\text{-}\lim_{k\to\infty} I_k = I$ in $\mathcal{X}$. By Lemma 6.1 and Theorem 6.2, it follows that the announced limits hold in the Polish spaces $\mathcal{P}_{\mathbb{R}^{2d}}$ and $\mathcal{P}_{\mathbb{R}^{d}}$. They also hold in $C'_{\mathbb{R}^{d}}$ and $C'_{\mathbb{R}^{2d}}$ since the effective domains of $S_k$ and $S$ and of $T_k$ and $T$ (considered as functions on $C'$) are included in $\mathcal{P}_{\mathbb{R}^{2d}}$ and $\mathcal{P}_{\mathbb{R}^{d}}$. This proves (a) and (b). Statement (c) follows from [15], Proposition 8.1. □

Let $\{g_i; i\ge 1\}$ be a countable subset of $C_{\mathbb{R}^d}$ such that $d(\gamma,\nu) = \sum_{i\ge1} 2^{-i}(|\langle g_i, \gamma-\nu\rangle| \wedge 1)$, $\gamma,\nu \in \mathcal{P}_{\mathbb{R}^d}$ is a metric which is compatible with the narrow convergence topology on $\mathcal{P}_{\mathbb{R}^d}$. For all $\nu \in \mathcal{P}_{\mathbb{R}^d}$ and all $\rho \in C'_{\mathbb{R}^{2d}}$, define

$$d(\rho_1,\nu) = \sum_{i\ge1} 2^{-i}(|\langle g_i\circ X_1, \rho\rangle - \langle g_i,\nu\rangle| \wedge 1).$$

Let us recall the three minimization problems

$$(\mathrm{MK}_k^\alpha) \qquad \text{minimize} \quad \frac{1}{k}H(\rho|\pi^k) + \alpha d(\rho_1,\nu) \quad \text{subject to} \quad \rho\in\Pi_0(\mu).$$

$$(\mathrm{MK}^\alpha) \qquad \text{minimize} \quad \int_{\mathbb{R}^{2d}} c\,d\rho + \alpha d(\rho_1,\nu) \quad \text{subject to} \quad \rho\in\Pi_0(\mu).$$

$$(\mathrm{MK}) \qquad \text{minimize} \quad \int_{\mathbb{R}^{2d}} c\,d\rho \quad \text{subject to} \quad \rho\in\Pi(\mu,\nu).$$

**Theorem 6.6.** *Assume that* $\mathcal{T}_c(\mu,\nu) < \infty$.

(a) *We have:* $\lim_{\alpha\to\infty}\lim_{k\to\infty}\inf_{\rho\in\Pi_0(\mu)}\left\{\frac{1}{k}H(\rho|\pi^k) + \alpha d(\rho_1,\nu)\right\} = \mathcal{T}_c(\mu,\nu)$.

(b) *For each* $k$ *and* $\alpha$, *(*$MK_k^\alpha$*) admits a unique solution* $\rho_k^\alpha$ *in* $\mathcal{P}_{\mathbb{R}^{2d}}$. *For each* $\alpha$, $(\rho_k^\alpha)_{k\ge1}$ *is a relatively compact sequence in* $\mathcal{P}_{\mathbb{R}^{2d}}$ *and any limit point of* $(\rho_k^\alpha)_{k\ge1}$ *is a solution of (*$MK^\alpha$*).*

(c) *For each* $\alpha$, *(*$MK^\alpha$*) admits at least a (possibly not unique) solution* $\rho^\alpha$. *The sequence* $(\rho^\alpha)_{\alpha\ge1}$ *is relatively compact in* $\mathcal{P}_{\mathbb{R}^{2d}}$ *and any limit point of* $(\rho^\alpha)_{\alpha\ge1}$ *is a solution of (MK).*

*Proof.* We introduce functions on $C'_{\mathbb{R}^{2d}}$ corresponding to $(\mathrm{MK}_k^\alpha)$, $(\mathrm{MK}^\alpha)$ and (MK). They are defined for all $\rho \in C'_{\mathbb{R}^{2d}}$ and each $k,\alpha\ge1$ by

$$\begin{aligned} G_k^\alpha(\rho) &= S_k(\rho) + \alpha d(\rho_1,\nu) = \frac{1}{k}H(\rho|\pi^k) + \xi_{\Pi_0(\mu)}(\rho) + \alpha d(\rho_1,\nu) \\ G^\alpha(\rho) &= S(\rho) + \alpha d(\rho_1,\nu) = \int_{\mathbb{R}^{2d}} c\,d\rho + \xi_{\Pi_0(\mu)}(\rho) + \alpha d(\rho_1,\nu) \\ G(\rho) &= \int_{\mathbb{R}^{2d}} c\,d\rho + \xi_{\Pi(\mu,\nu)}(\rho). \end{aligned}$$

The domains of $S_k$ and $S$ are included in the strong unit ball $U_{\mathbb{R}^{2d}}$ of $C'_{\mathbb{R}^{2d}}$. Therefore, the domains of $G_k^\alpha, G_k$ and $G$ are also in $U_{\mathbb{R}^{2d}}$ which is $\sigma(C'_{\mathbb{R}^{2d}}, C_{\mathbb{R}^{2d}})$-compact.
We know that $S_k, S$ are lower semicontinuous, $d(\rho_1,\nu)$ is continuous and bounded below

and $\Pi_1(\nu)$ is closed. Therefore, $G_k^\alpha, G^\alpha$ and $G$ are inf-compact.
As the relative entropy is stricly convex, $G_k^\alpha$ is also strictly convex: it admits a unique minimizer $\rho_k^\alpha$.

As a function of $\rho$, $d(\rho_1,\nu)$ is a finite continuous function on $C'_{\mathbb{R}^{2d}}$. Together with the convergence $\Gamma\text{-}\lim_{k\to\infty} S_k = S$, this implies (see [15], Proposition 6.21) that for all $\alpha$,

$$\Gamma\text{-}\lim_{k\to\infty} G_k^\alpha = G^\alpha \quad \text{in } C'_{\mathbb{R}^{2d}}.$$

Observe that $\lim_{\alpha\to\infty} \alpha d(\rho_1,\nu) = \xi_{\Pi_1(\nu)}(\rho)$ for all $\rho \in \mathcal{P}_{\mathbb{R}^{2d}}$. As this limit is increasing, by [15], Proposition 5.4 we have

$$\Gamma\text{-}\lim_{\alpha\to\infty} G^\alpha = G \quad \text{in } C'_{\mathbb{R}^{2d}}.$$

Together with the relative compactness of the domains, these $\Gamma$-convergence results entail the whole theorem (see [15], Theorem 7.8 and Corollary 7.20). ☐

## 7. Γ-convergence of convex functions on a weakly compact space

This section is dedicated to the proof of Corollary 7.4 which is an important tool for the proof of Theorem 4.9.

A typical result about the $\Gamma$-convergence of a sequence of convex functions $(f_n)$ is: If the sequence of the convex conjugates $(f_n^*)$ converges in some sense, then $(f_n)$ $\Gamma$-converges. Known results of this type are usually stated in separable reflexive Banach spaces. For instance Corollary 3.13 of H. Attouch's monograph [1] is

**Theorem 7.1.** *Let $X$ be a separable reflexive Banach space and $(f_n)$ a sequence of closed convex functions from $X$ into $(-\infty,+\infty]$ satisfying the equicoerciveness assumption: $f_n(x) \geq \alpha(\|x\|)$ for all $x \in X$ and $n \geq 1$ with $\lim_{r\to+\infty} \alpha(r)/r = +\infty$. Then, the following statements are equivalent*

(1) $f = \mathrm{seq}X_w\text{-}\Gamma\text{-}\lim_{n\to\infty} f_n$
(2) $f^* = X_s^*\text{-}\Gamma\text{-}\lim_{n\to\infty} f_n^*$
(3) $\forall y \in X^*,\ f^*(y) = \lim_{n\to\infty} f_n^*(y)$

*where $X^*$ is the dual space of $X$, $\mathrm{seq}X_w$ refers to the weak sequential convergence in $X$ and $X_s^*$ to the strong convergence in $X^*$.*

Escaping from the reflexivity assumption is quite difficult, as can be seen in G. Beer's monograph [2].

In some applications in probability, the reflexive Banach space setting is not as natural as it is for the usual applications of variational convergence to PDEs. For instance when dealing with random measures on $\mathcal{X}$, the narrow topology $\sigma(\mathcal{P}_\mathcal{X}, C_b(\mathcal{X}))$ doesn't fit the above framework since $C_b(\mathcal{X})$ endowed with the uniform topology may not be separable (unless $\mathcal{X}$ is compact) and is not reflexive.

The next result is an analogue of Theorem 7.1 which agrees with applications for random probability measures. Since we didn't find it in the literature, we give its detailed proof.

Let $X$ and $Y$ be two vector spaces in separating duality. The space $X$ is furnished with the weak topology $\sigma(X,Y)$.

We denote $\xi_C$ the indicator function of the subset $C$ of $X$ which is defined by $\xi_C(x) = 0$ if $x$ belongs to $C$ and $\xi_C(x) = +\infty$ otherwise. Its convex conjugate is the support function of $C$ : $\xi_C^*(y) = \sup_{x\in C}\langle x,y\rangle$, $y \in Y$.

**Theorem 7.2.** *Let $(g_n)$ be a sequence of functions on $Y$ such that*

(a) *for all $n$, $g_n$ is a real-valued convex function on $Y$,*

(b) *$(g_n)$ converges pointwise to $g := \lim_{n\to\infty} g_n$,*
(c) *$g$ is real-valued and*
(d) *in restriction to any finite dimensional vector subspace $Z$ of $Y$, $(g_n)$ $\Gamma$-converges to $g$, i.e. $\Gamma\text{-}\lim_{n\to\infty}(g_n + \xi_Z) = g + \xi_Z$, where $\xi_Z$ is the indicator function of $Z$.*

*Denote the convex conjugates on $X$ : $f_n = g_n^*$ and $f = g^*$.*

*If in addition,*

(e) *there exists a compact set $K \subset X$ such that $\operatorname{dom} f_n \subset K$ for all $n \geq 1$ and $\operatorname{dom} f \subset K$*

*then, $(f_n)$ $\Gamma$-converges to $f$ with respect to $\sigma(X,Y)$.*

*Remark* 7.3. By ([15], Proposition 5.12), under the assumption (a), assumption (d) is implied by:

(d') in restriction to any finite dimensional vector subspace $Z$ of $Y$, $(g_n)$ is equibounded, i.e. for all $y_o \in Z$, there exists $\delta > 0$ such that

$$\sup_{n\geq 1} \sup\{|g_n(y)|; y \in Z, |y - y_o| \leq \delta\} < \infty.$$

A useful consequence of Theorem 7.2 is

**Corollary 7.4.** *Let $(Y, \|\cdot\|)$ be a normed space and $X$ its topological dual space. Let $(g_n)$ be a sequence of functions on $Y$ such that*

(a) *for all $n$, $g_n$ is a real-valued convex function on $Y$,*
(b) *$(g_n)$ converges pointwise to $g := \lim_{n\to\infty} g_n$ and*
(d") *there exists $c > 0$ such that $|g_n(y)| \leq c(1 + \|y\|)$ for all $y \in Y$ and $n \geq 1$.*

*Then, $(f_n)$ $\Gamma$-converges to $f$ with respect to $\sigma(X,Y)$ where $f_n = g_n^*$ and $f = g^*$.*

*Proof.* Under (b), (d") implies (c). Since the functions $g_n$ are convex, (d") implies that $\{g_n; n \geq 1\}$ is locally equi-Lipschitz. Therefore (d") implies (d') and we have (d) by Remark 7.3. Finally, (d") implies (e) with $K = \{x \in X; \|x\|_* \leq c\}$ where $\|x\|_* = \sup_{y,\|y\|\leq 1}\langle x, y\rangle$ is the dual norm on $X$. Indeed, suppose that for all $y \in Y$, $g(y) \leq c + c\|y\|$ and take $x \in X$ such that $g^*(x) < +\infty$. As for all $y$, $\langle x, y\rangle \leq g(y) + g^*(x)$, we get $|\langle x, y\rangle|/\|y\| \leq (g^*(x) + c)/\|y\| + c$. Letting $\|y\|$ tend to infinity gives $\|x\|_* \leq c$ which is the announced result.
The conclusion follows from Theorem 7.2. □

The proof of Theorem 7.2 is postponed after the two preliminary Lemmas 7.5 and 7.11.

**Lemma 7.5.** *Let $f : X \to (-\infty, +\infty]$ be a lower semicontinuous convex function such that $\operatorname{dom} f$ is included in a compact set. Let $V$ be a closed convex subset of $X$.*

*Then, if $V$ satisfies*

$$V \cap \operatorname{dom} f \neq \emptyset \quad \text{or} \quad V \cap \operatorname{cl}\operatorname{dom} f = \emptyset, \tag{7.6}$$

*we have*

$$\inf_{x\in V} f(x) = -\inf_{y\in Y}(f^*(y) + \xi_V^*(-y)) \in (-\infty, \infty] \tag{7.7}$$

*and if $V$ doesn't satisfy (7.6), we have*

$$\inf_{x\in W} f(x) = -\inf_{y\in Y}(f^*(y) + \xi_W^*(-y)) = +\infty \tag{7.8}$$

*for all closed convex set $W$ such that $W \subset \operatorname{int} V$.*

*Proof.* The proof is divided in two parts. We first consider the case where $V \cap \mathrm{dom}\, f \neq \emptyset$, then the case where $V \cap \mathrm{cl}\,\mathrm{dom}\, f = \emptyset$.

• *The case where* $V \cap \mathrm{dom}\, f \neq \emptyset$. As $V$ is a nonempty closed convex set, its indicator function $\xi_V$ is a closed convex function so that its biconjugate satisfies $\xi_V^{**} = \xi_V$, i.e. $\xi_V(x) = \sup_{y\in Y}\{\langle x,y\rangle - \xi_V^*(y)\}$ for all $x \in X$. Consequently,

$$\inf_{x\in V} f(x) = \inf_{x\in X}\sup_{y\in Y}\{f(x) + \langle x,y\rangle - \xi_V^*(y)\}.$$

One wishes to invert $\inf_{x\in X}$ and $\sup_{y\in Y}$ by means of the following standard inf-sup theorem (see [9] for instance). We have $\inf_{x\in X}\sup_{y\in Y} F(x,y) = \sup_{y\in Y}\inf_{x\in X} F(x,y)$ provided that $\inf_{x\in X}\sup_{y\in Y} F(x,y) \neq \pm\infty$ and

- $\mathrm{dom}\, F$ is a product of convex sets,
- $x \mapsto F(x,y)$ is convex and lower semicontinuous for all $y$,
- there exists $y_o$ such that $x \mapsto F(x,y_o)$ is inf-compact and
- $y \mapsto F(x,y)$ is concave for all $x$.

Our assumptions on $f$ allow us to apply this result with $F(x,y) = f(x) + \langle x,y\rangle - \xi_V^*(y)$. Note that

$$\inf_{x\in X} f(x) > -\infty \tag{7.9}$$

since $f$ doesn't take the value $-\infty$ and is assumed to be lower semicontinuous on a compact set. Therefore, if $\inf_{x\in V} f(x) < +\infty$, we have

$$\inf_{x\in V} f(x) = \sup_{y\in Y}\inf_{x\in X}\{f(x) + \langle x,y\rangle - \xi_V^*(y)\} = -\inf_{y\in Y}\{f^*(y) + \xi_V^*(-y)\}.$$

• *The case where* $V \cap \mathrm{cl}\,\mathrm{dom}\, f = \emptyset$. As $\mathrm{cl}\,\mathrm{dom}\, f$ is assumed to be compact, by Hahn-Banach theorem $\mathrm{cl}\,\mathrm{dom}\, f$ and $V$ are strictly separated: there exists $y_o \in Y$ such that $\xi_V^*(y_o) = \sup_{x\in V}\langle x,y_o\rangle < \inf_{\mathrm{cl}\,\mathrm{dom}\, f}\langle x,y_o\rangle \le \inf_{x\in\mathrm{dom}\, f}\langle x,y_o\rangle$. Hence,

$$\inf_{x\in\mathrm{dom}\, f}\{\langle x,y_o\rangle - \xi_V^*(y_o)\} > 0 \tag{7.10}$$

and

$$\begin{aligned}
-\inf_{y\in Y}(f^*(y) + \xi_V^*(-y)) &= \sup_{y\in Y}\inf_{x\in X}\{f(x) + \langle x,y\rangle - \xi_V(y)\}\\
&= \sup_{y\in Y}\inf_{x\in\mathrm{dom}\, f}\{f(x) + \langle x,y\rangle - \xi_V(y)\}\\
&\ge \inf_{x\in X} f(x) + \sup_{a>0}\inf_{x\in\mathrm{dom}\, f}\{\langle x,ay_o\rangle - \xi_V^*(ay_o)\}\\
&= \inf_{x\in X} f(x) + \sup_{a>0} a \inf_{x\in\mathrm{dom}\, f}\{\langle x,y_o\rangle - \xi_V^*(y_o)\}\\
&= +\infty
\end{aligned}$$

where the last equality follows from (7.9) and (7.10). This proves that (7.8) holds with $W = V$.

• Finally, if (7.6) isn't satisfied, taking $W$ such that $W \subset \mathrm{int}\, V$ insures the strict separation of $W$ and $\mathrm{cl}\,\mathrm{dom}\, f$ as above. □

**Lemma 7.11.** *Let the* $\sigma(X,Y)$*-closed convex neighbourhood* $V$ *of the origin be defined by*

$$V = \{x \in X; \langle y_i, x\rangle \le 1, 1 \le i \le k\}$$

*with* $k \ge 1$ *and* $y_1,\dots,y_k \in Y$. *Its support function* $\xi_V^*$ *is* $[0,\infty]$*-valued, inf-compact and its domain is the finite dimensional convex cone spanned by* $\{y_1,\dots,y_k\}$. *More precisely, its level sets are* $\{\xi_V^* \le b\} = b\,\mathrm{cv}\{y_1,\dots,y_k\}$ *for each* $b \ge 0$ *where* $\mathrm{cv}\{y_1,\dots,y_k\}$ *is the convex hull of* $\{y_1,\dots,y_k\}$.

*Proof.* The closed convex set $V$ is the polar set of $N = \{y_1, \dots, y_k\} : V = N^\circ$. Let $x_1 \in V$ and $x_o \in E := \cap_{1\le i\le k} \ker y_i$. Then, $\langle y_i, x_1 + x_o\rangle = \langle y_i, x_1\rangle \le 1$. Hence, $x_1 + x_o \in V$. Considering the factor space $X/E$, we now work within a finite dimensional vector space whose algebraic dual space is spanned by $\{y_1, \dots, y_k\}$.

We still denote by $X$ and $Y$ these finite dimensional spaces. We are allowed to apply the finite dimension results which are proved in the book [22] by Rockafellar and Wets. In particular, one knows that if $C$ is a closed convex set in $Y$, then the gauge function $\gamma_C(y) := \inf\{\lambda \ge 0; y \in \lambda C\}, y \in Y$ is the support function of its polar set $C^\circ = \{x \in X; \langle x, y\rangle \le 1, \forall y \in C\}$. This means that $\gamma_C = \xi^*_{C^\circ}$ (see [22], Example 11.19).

As $V = (N^{\circ\circ})^\circ$ and $N^{\circ\circ}$ is the closed convex hull of $N$, i.e. $N^{\circ\circ} = \mathrm{cv}(N)$ : the convex hull of $N$, we get $V = \mathrm{cv}(N)^\circ$ and

$$\xi^*_V = \gamma_{\mathrm{cv}(N)}.$$

In particular, for all real $b$, $\xi^*_V(y) \le b \Leftrightarrow \gamma_{\mathrm{cv}(N)}(y) \le b \Leftrightarrow y \in b\,\mathrm{cv}(N)$. It follows that the effective domain of $\xi^*_V$ is the convex cone spanned by $y_1, \dots, y_k$ and $\xi^*_V$ is inf-compact. $\square$

*Proof of Theorem 7.2.* Let $\mathcal{N}(x_o)$ denote the set of all the neighbourhoods of $x_o \in X$. We want to prove that $\Gamma\text{-}\lim_{n\to\infty} f_n(x_o) := \sup_{U\in\mathcal{N}(x_o)} \lim_{n\to\infty} \inf_{x\in U} f_n(x) = f(x_o)$. Since $f$ is lower semicontinuous, we have $f(x_o) = \sup_{U\in\mathcal{N}(x_o)} \inf_{x\in U} f(x)$, so that it is enough to show that for all $U \in \mathcal{N}(x_o)$, there exists $V \in \mathcal{N}(x_o)$ such that $V \subset U$ and

$$\lim_{n\to\infty} \inf_{x\in V} f_n(x) = \inf_{x\in V} f(x). \tag{7.12}$$

The topology $\sigma(X, Y)$ is such that $\mathcal{N}(x_o)$ admits the sets

$$V = \{x \in X; |\langle y_i, x - x_o\rangle| \le 1, i \le k\}$$

as a base where $(y_1, \dots, y_k), k \ge 1$ describes the collection of all the finite families of vectors in $Y$. By Lemma 7.5, there exists such a $V \subset U$ which satisfies

$$\inf_{x\in V} f_n(x) = -\inf_{y\in Y} h_n(y) \text{ for all } n \ge 1 \text{ and } \inf_{x\in V} f(x) = -\inf_{y\in Y} h(y)$$

where we denote $h_n(y) = g_n(y) + \xi^*_V(-y)$ and $h(y) = g(y) + \xi^*_V(-y)$, $y \in Y$.

Let $Z$ denote the vector space spanned by $(y_1, \dots, y_k)$ and $h^Z_n, h^Z$ the restrictions to $Z$ of $h_n$ and $h$. For all $y \in Y$, we have

$$\xi^*_V(-y) = -\langle x_o, y\rangle + \xi^*_{V-x_o}(-y) \tag{7.13}$$

and by Lemma 7.11, the effective domain of $\xi^*_V$ is $Z$. Therefore, to prove (7.12) it remains to show that

$$\lim_{n\to\infty} \inf_{y\in Y} h^Z_n(y) = \inf_{y\in Y} h^Z(y). \tag{7.14}$$

By assumptions (b) and (d), $(h^Z_n)$ $\Gamma$-converges and pointwise converges to $h^Z$. Note that this $\Gamma$-convergence is a consequence of the lower semicontinuity of the convex conjugate $\xi^*_V$ and Proposition 6.25 of [15].

Because of assumptions (a) and (c), $(h^Z_n)$ is also a sequence of finite convex functions which converges pointwise to the finite function $h^Z$. By ([21], Theorem 10.8), $(h^Z_n)$ converges to $h^Z$ uniformly on any compact subset of $Z$ and $h^Z$ is convex.

We now consider three cases for $x_o$.

*The case where $x_o \in \mathrm{dom}\, f$.* We already know that $(h^Z_n)$ $\Gamma$-converges to $h^Z$. To prove (7.14), it remains to check that the sequence $(h^Z_n)$ is equicoercive (see [15], Theorem 7.8). For all $y \in Y$, $g(y) - \langle x_o, y\rangle \ge -f(x_o)$ and (7.13) imply $h^Z(y) \ge -f(x_o) + \xi^*_{V-x_o}(-y)$.

Since, $-f(x_o) > -\infty$ and $\xi^*_{V-x_o}$ is inf-compact (Lemma 7.11), we obtain that $h^Z$ is inf-compact. As $(h^Z_n)$ converges to $h^Z$ uniformly on any compact subset of $Z$, it follows that $(h^Z_n)$ is equicoercive. This proves (7.14).

*The case where* $x_o \in \mathrm{cl}\,\mathrm{dom}\, f$. In this case, there exists $x'_o \in \mathrm{dom}\, f$ such that $V' = x'_o + (V - x_o)/2 = \{x \in X; |\langle 2y_i, x - x'_o\rangle| \le 1, i \le k\} \in \mathcal{N}(x'_o)$ satisfies $x_o \in V' \subset V \subset U$. One deduces from the previous case, that (7.14) holds true with $V'$ instead of $V$.

*The case where* $x_o \notin \mathrm{cl}\,\mathrm{dom}\, f$. As $(h^Z_n)$ $\Gamma$-converges to $h^Z$, by ([2], Proposition 1.3.5) we have $\limsup_{n\to\infty} \inf_{y\in Y} h^Z_n(y) \le \inf_{y\in Y} h^Z(y)$. As $x_o \notin \mathrm{cl}\,\mathrm{dom}\, f$, for any small enough $V \in \mathcal{N}(x_o)$, $\inf_{y\in Y} h^Z(y) = -\inf_{x\in V} f(x) = -\infty$. Therefore, $\lim_{n\to\infty} \inf_{y\in Y} h^Z_n(y) = \inf_{y\in Y} h(y) = -\infty$ which is (7.14).

This completes the proof of Theorem 7.2. □

## References


[1] H. Attouch. *Variational convergence for functions and operators.* Pitman Advanced Publishing Program. Pitman, 1984.

[2] G. Beer. *Topologies on closed and closed convex sets*, volume 268 of *Mathematics and Its Applications.* Kluwer Academic Publishers, 1993.

[3] C. Boucher, R.S. Ellis, and B. Turkington. Spatializing random measures: doubly indexed processes and the large deviation principle. *Ann. Probab.*, 27:297–324, 1999.

[4] A. Braides. *Γ-convergence for Beginners.* Oxford Lecture Series in Mathematics 22. Oxford University Press, 2002.

[5] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C.R. Acad. Sci. Paris, Série I*, 305:805–808, 1987.

[6] P. Cattiaux and C. Léonard. Large deviations and Nelson's processes. *Forum Math.*, 7:95–115, 1995.

[7] D. A. Dawson and J. Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20:247–308, 1987.

[8] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications. Second edition.* Applications of Mathematics 38. Springer Verlag, 1998.

[9] I. Ekeland. *La théorie des jeux et ses applications à l'économie mathématique.* Presses Universitaires de France, 1974.

[10] D. Feyel and A. S. Üstünel. Monge-Kantorovitch measure transportation and Monge-Ampère equation on Wiener space. *Probab. Theory Related Fields*, 128(3):347–385, 2004.

[11] D. Feyel and A. S. Üstünel. Monge-Kantorovitch measure transportation, Monge-Ampère equation and the Itô calculus. In *Stochastic analysis and related topics in Kyoto*, volume 41 of *Adv. Stud. Pure Math. Math. Soc. Japan*, pages 49–74, Tokyo, 2004.

[12] L. V. Kantorovich. On the translocation of masses. *C. R. (Dokl.) Acad. Sci. URSS*, 37:199–201, 1942.

[13] L. V. Kantorovich. On a problem of Monge (in Russian). *Uspekhi Mat. Nauk.*, 3:225–226, 1948.

[14] C. Léonard. Large deviations of doubly indexed systems. Preprint, 2005.

[15] G. Dal Maso. *An Introduction to Γ-Convergence.* Progress in Nonlinear Differential Equations and Their Applications 8. Birkhäuser, 1993.

[16] T. Mikami. Monge's problem with a quadratic cost by the zero-noise limit of $h$-path processes. *Probab. Theory Relat. Fields*, 129:245–260, 2004.

[17] G. Monge. Mémoire sur la théorie des déblais et des remblais. In *Histoire de l'Académie Royale des Sciences de Paris*, pages 666–704. 1781.

[18] J. Neveu. *Bases mathématiques du calcul des probabilités.* Masson, Paris, 1970.

[19] S. Rachev and L. Rüschendorf. *Mass Transportation Problems. Vol I : Theory, Vol. II : Applications.* Probability and its applications. Springer Verlag, New York, 1998.

[20] R. T. Rockafellar. Convex integral functionals and duality. In E. Zarantonello, editor, *Contributions to nonlinear functional analysis*, pages 215–235. Academic Press, New-York, 1971.

[21] R.T. Rockafellar. *Convex Analysis.* Princeton landmarks in mathematics. Princeton University Press, Princeton, N.J., 1997. First published in the Princeton Mathematical Series in 1970.

[22] R.T. Rockafellar and R. Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften.* Springer, 1998.

[23] D. Ruelle. *Thermodynamic Formalism.* Addison Wesley, Reading, MA, 1978.
[24] C. Villani. *Topics in Optimal Transportation.* Graduate Studies in Mathematics 58. American Mathematical Society, Providence RI, 2003.
[25] C. Villani. Saint-Flour Lecture Notes. Optimal transport, old and new.
Available online via `http://www.umpa.ens-lyon.fr/~cvillani/`, 2005.

(Christian Léonard) Modal-X, Université Paris 10. Bât. G, 200 av. de la République. 92001 Nanterre Cedex, France

(Christian Léonard) CMAP, École Polytechnique. 91128 Palaiseau Cedex, France
*E-mail address*: `christian.leonard@polytechnique.fr`